Systems for analysing and assessing language learning in the classroom

Harrison, Andrew

Systems for analysing and assessing language learning in the classroom

Andrew Harrison

PhD Thesis
King's College, London
2000

Abstract

The starting point for the thesis is the problem of producing classroom assessment materials which are compatible with current language learning methodology. An account of language learning for communicative purposes is followed by an exploration of how far testing has reflected the same interests, an exercise which reveals a mismatch between teaching materials and assessment procedures. To establish what, in practice, communicative learning materials demand of students, a system is developed for the analysis of the content of course books. This provides a basis for assessments which reflect the materials and methods used in a particular classroom, taking into account the students' current contexts, both social (as individuals and groups within the school) and geographical (relating to the location of the school and the likely contacts of students). The thesis then reports on a project, undertaken with the collaboration of teachers and learners, which aims to put into effect the proposed sequence of an analysis system followed by an assessment system.
Finally, the project is reviewed and evaluated.

Contents

Figures

Tables

Acknowledgements

I am very grateful to all those who have helped in so many different ways with the work for this thesis and whose support has made it possible, and especially Nick Andon, Tom Bradbury, Sara Bunnell, Rose Clark, Ray Edwards, Pauline Foster, Michael Garman, Miyoko Kobayashi, Henry Macintosh, Andy Mackay, Terry Miles, Dermot Murphy, Matthew Portal, Pauline Rea-Dickins, Pippa Stone, and Guy Wellman. I am also indebted to the teachers who helped me work through my ideas with them and the students who took part in the trialling, for without them the ideas would have remained abstract and untried.

My greatest debts are however to my family, particularly Mary and Mark, and above all to my supervisor, Peter Skehan, who has given me wise advice and unflagging support over many years, particularly at the bleakest moments. Without his insistence that the writing should represent the work done I should never have achieved this much.

All errors and omissions are of course my own responsibility.

Chapter one

A communicative approach: theory & practice

1.1 Starting points

The original impetus for this thesis was the view that language learning and language testing should be more closely related. The adoption of a communicative approach to teaching and learning has not been followed by a commensurate change in testing and assessment methods, but there are two implementations of communicative theories which suggest that the relationship between learning and assessment is crucial: they are useful starting points for the discussion which follows in the first two chapters, and underlie the project which is reported on later.

The first Modern Languages project of the Council for Cultural Co-operation of the Council of Europe, starting from a conference at Rüschlikon, Switzerland in 1971 was a vast undertaking intended to 'make the free movement of men [sic] and ideas in the European area easier by increasing the scale and effectiveness of language learning' (van Ek 1975). The Graded Objectives in Modern Languages movement (which came to be known as GOML) was a series of local initiatives by teachers supported by Local Education Authorities in the UK as a means of motivating the new, unselected secondary school learners of French and German in secondary comprehensive schools (Page 1973 and 1974, Davidson 1973). Both developments are examples of a communicative approach to language learning, but the differences between the two are relevant to the theme of this thesis. The Council of Europe were primarily interested in language learning and its social impact over a large number of different countries - Trim (1996: 81) reports that by 1996 there were 44 states involved - whereas the British developments were a series of local responses to the lack of materials available for teaching pupils of lower ability and the likelihood that these pupils would not be adequately served by the external examination system then existing. The Council of Europe project was interested first in language learning for adults, and the need for relevant tests was discussed subsequently (although implied from the start in the idea of a continent-wide 'Unit/Credit System' - Trim et al 1973); GOML schemes in general began with ideas for new kinds of test and then (in most schemes) undertook the design of materials to prepare for them (Harrison 1985). In both cases, teaching and learning on the one hand and testing and assessment on the other were seen as interdependent means of fulfilling the same end: the learning of applicable language skills.

The underlying motive of this thesis is the view that testing, or more broadly assessment, should be more frequently considered as part of learning and should

therefore be much more integrated with classroom events than it is conventionally allowed to be. As in the Council of Europe's approach, the classroom should be preparing learners for active use of the language in the future; as with the GOML schemes, assessment should relate directly to what has preceded it by reflecting the aims and content of an applications-oriented methodology. In addition, assessment should occur regularly in the classroom, and relatively informally, so that it helps learners to appreciate what they can achieve in cooperation with each other and with the teacher. The natural consequence of this argument is that it is the teacher who should be undertaking the assessment of his* students, preferably with the help of colleagues, and especially if some formality is to be attached to the results.

1.2 Background

The 'communicative approach' to language learning began in the early 1970s as a general movement towards change in second language (L2) teaching methods which seemed revolutionary at the time, though it contained elements of previous concepts of how foreign languages are to be learnt. Howatt (1984) sees various aspects of communicative ideas in movements given such labels as Natural Method, Conversation Method and Direct Method, and usefully sums up the underlying philosophy.

> Learning how to speak a new language, it is held, is not a rational process which can be organized in a step-by-step manner following graded syllabuses of new points to learn, exercises and explanation. It is an intuitive process for which human beings have a natural capacity that can be awakened provided only that the proper conditions exist. ...someone to talk to, something to talk about and a desire to understand and make yourself understood.
>
> Howatt 1984:192

Brumfit & Johnson (1979:24) suggest three areas contributing to communicative teaching: the sociological (sociolinguistic), the philosophical and the linguistic. In a later paper, Brumfit defines in short what the term 'communicative' implies (without however using it):

> The central issue is the insistence that language is fluid, dynamic and negotiable and that this fact needs to be recognized in the process of teaching.
>
> Brumfit 1984:314

He then goes on to list six sources for this view of language: linguistic discussion and the notion of communicative competence; anthropology, with speech events influenced by, for example, participants, setting, topic; sociolinguistics and concern with social context; social psychology with its recognition of in- and out-groups; philosophy and the concept of the speech act; and ethnomethodology, which includes emphasis on negotiated conventions and the systematic nature of much apparently spontaneous activity (1984:314-5). (For a full account of these and other more remote influences on current thinking, see Howatt 1984.)

As the arguments for communicative methods developed, a variety of explorations of the field covered the ground in some detail from different points of view (starting

* Throughout this thesis, the masculine form is used without intended prejudice to represent both male and female genders.

perhaps with Wilkins 1972 and 1974, Widdowson 1978 and 1979 and Brumfit & Johnson 1979, following up with Johnson & Morrow (eds) 1981, Johnson 1982, Brumfit (ed) 1986, Clarke 1987, with discussions multiplying over the years up to Cook & Seidlhofer (eds) 1995, and still continuing).

1.3 'Communicative' in theory

A useful way forward at this point in the present discussion could be to attempt a distillation from key texts which set out some of the principles of communicative language teaching and to draw from it guidelines which may eventually become a scaffolding to support the building of the present argument. Morrow (1981), for example, suggests five principles which 'might guide us in search of a [communicative] method':

1. Know what you are doing ('every lesson should be an operation of some kind which the student might actually want to perform in the foreign language');

2. The whole is more than the sum of the parts ('a communicative method ... operates with stretches of language above the sentence level, and operates with real language in real situations');

3. The processes are as important as the forms, eg - information gap ('one student must be in a position to tell another something that the second student does not already know') - choice ('the speaker ... must choose not only what ideas he wants to express at a given moment, but also what linguistic forms are appropriate to express them') - feedback ('what you say to somebody depends not only on what he has just said to you (though this is obviously very important) but also on what you want to get out of the conversation');

4. To learn it, do it ('what happens in the classroom must involve the learner and must be judged in terms of its effects on him'); and

5. A mistake is not always a mistake ('a communicative method ... may well ...require the flexibility to treat different things as "mistakes" at different stages in the learning process') (Morrow: 1981 59-66).

Johnson's (1982:163ff) 'five principles as demonstrated in a "communicative" exercise type' are: information transfer, information gap, jigsaw, task dependency and correction for content.

1. 'Information transfer' is concerned with asking the student to extract information and use it for a purpose.

2. An 'information gap' occurs when there is 'genuine information flow in the class; the students tell each other things they do not already know.'

3. The 'jigsaw' principle (adapted here by Johnson from an idea first exploited by Geddes & Sturtridge (1979) for listening exercises) requires that students should work at tasks they cannot complete satisfactorily without the contribution of others.

4. 'Task dependency' is explained as the need for the student to use the information given in some way for a reason, if only his accountability to another student for comprehensible interchange.

5. The 'correction for content' principle measures up a student's language production in two ways, for communicative efficacy and grammatical accuracy, the latter *at some other stage* (Johnson's italics).

It is apparent that these two early leaders in communicative methodology differ in what they see as the most important principles of the approach even though they worked together as colleagues (eg Johnson & Morrow 1979, Morrow & Johnson 1979). They agree that students should use information for a reason or to achieve something they might want to perform outside the classroom; that they should be telling each other things they don't already know; that they should be using language purposefully; and that errors should be treated differently depending on what kind of exercise the students are engaged in. But Morrow also gives priority to realism, choice and learning by doing, and Johnson includes cooperation. These principles are now to be explored further under the headings *application, information gap, purpose, error, realism, choice, learning by doing, cooperation.*

Principle 1

Application: The students should use information for a reason or to achieve something they might want to perform outside the classroom.

This principle opens up the whole area of needs: the point of learning in a communicative approach is that it should be a direct preparation for communicating with others. But there is a problem with this apparently logical idea, which Wilkins puts succinctly:

> Most teachers, working within school systems, face a situation where accurate prediction of future language needs is scarcely possible. In that case the teacher has to make certain assumptions about what will be of greatest general value to his pupils.
>
> Wilkins 1974:59

Brumfit agrees that the teacher does not often have the opportunity to take into account the needs of his students: he has to 'teach as efficiently as possible within the social and administrative constraints imposed'. Brumfit sees effective communicative strategies as providing 'one major way of maintaining motivation' (Brumfit 1981b:47). A means of engaging this motivation is to ask secondary school pupils themselves what they would like to see included in their foreign language

courses (Clark 1987:144). Those working with adults, particularly in contexts where Special Purposes can be identified, have an easier task in shaping learning to practical ends. The most comprehensive attempt at a description of needs in this sense is probably that developed by Munby (1978), but as a system it is at once too detailed and not individual enough. It is not applicable in wide areas of language learning (Nunan 1988:20), runs the risks of trivialisation of behaviour and a check-list approach (White 1983:76) and is 'ultimately unworkable' (Morrow 1983).

But the Council of Europe project, which began in 1971 and has continued to develop since (see Brumfit (ed) 1995), seems to have successfully reconciled broad aims and a restricted syllabus (eg van Ek 1975). Initially, it had a diverse clientèle in mind: adults learning languages in order to interchange information and ideas with others in an increasingly mobile population within Europe (Richterich & Chancerel 1978). These learners include both native speakers of various European languages and immigrant workers, and so are diverse within a common culture, or at least a culture to which the learners are to adjust. At the same time, it produced detailed syllabuses at defined levels, *Threshold* and *Waystage*, in several languages (Shiels 1995). For a less specific audience, whoever the learners may be, Ellis includes in his suggestions for an optimal communicative environment for classroom language learning, that 'the students have the opportunity to participate in the kind(s) of discourse (planned and/or unplanned) which correspond to their communicative needs outside the classroom' (Ellis 1990:127). And Widdowson suggests, with an even broader perspective: 'what seems to be needed is a presentation which sets problems... which are situation and not language centred... The crucial point is that they should provide for the development of abilities for coping with real life subsequently.' (Widdowson 1984:225).

The needs discussion should therefore be centred not on the words essential to topics, or the structures required for a dynamic, or the functions/notions/situations which may or may not confront the learner (all necessary on occasion, but incidental), nor yet on a system like Munby's which attempts a comprehensive but in the end impractical analysis of instrumental demands. It logically starts from the learner, as both Munby and Richterich & Chancerel insist, but ends with coping, with the possession of those skills which will enable the learner to deal with his future encounters with others and his ensuing entanglements with the language. (For a detailed account of needs analysis, see West 1994.)

Principle 2

Information gap: The students should be telling each other things they don't already know.

This principle may be appropriate for learners, but does not necessarily apply in the same way to talk from the teacher. Howatt quotes the principle laid down by Sauveur in 1874 that the teacher should ask only 'earnest' questions, that is, not necessarily ones to which he does not know the answer, but ones to which he is

genuinely looking for an answer.

> There is a view at the present time that the only genuine classroom questions are ones to which the teacher does not know the answer... all other 'questions' are merely code-practising devices. This is not, it seems to me, necessarily the case.
>
> Howatt 1984:201

Widdowson's (1978) simple but exact illustration of the dichotomy between a real question and a false one is to differentiate between teacher sentences such as 'what is on the table?' and 'where's the duster?', the answer to the former being obvious, but to the latter unknown to the questioner. The information gap principle thus starts at the level of a basic interchange in the classroom - question and answer. It rests on the distinction between what Brock (1986) calls 'display' questions (in answer to which the student shows that he knows something of the language) and 'reference' questions (where in his answer the student explains something he knows but others don't). But the implications vary depending on who are the contributors to the exchange. Display questions may serve a useful function in exchanges between teacher and learner, but are less likely to be realistic between learner and learner if the context is meant to be 'communicative.' The distinction between knowing and not knowing can be further developed through variations - from small-scale, limited ones like 'one speaker must not know what the other speaker is going to say' (Ellis 1982:75) to elaborated ones like Prabhu's three types of 'meaning-focused activity', information-gap, reasoning-gap and opinion-gap (Prabhu 1987:46). These all refer to learner-learner exchanges, with the assumption that there must be some communication to be made, otherwise the deployment of language is pointless. But Widdowson's example is only a small peg on which to hang a large argument, which is the distinction between 'usage' and 'use', a theme to which he constantly returns in later papers.

> The teaching of usage does not appear to guarantee a knowledge of use. The teaching of use, however, does seem to guarantee the learning of usage since the latter is represented as a necessary part of the former.
>
> Widdowson 1978:19

Allwright proposes a similar relationship in suggesting that linguistic competence is part of communicative competence.

> If this way of specifying the relationship is generally correct, ... we would be well advised to focus on communicative skills, in the knowledge that this will necessarily involve developing most areas of linguistic competence as an essential part of the product rather than focus on linguistic skills and risk failing to deal with a major part of whatever constitutes communicative competence.
>
> Allwright 1979:168

There are links between these concepts and the use of the necessary language for completing a task, including finding fitting solutions with the cooperation of a partner: in this case, the information gap lies between the language users and what they will need to do, since at the start, neither knows what the solution will be or

how they will get there.

Principle 3

Purpose: The students should be using the language purposefully.

There are several strands to 'purpose'. One of these is consideration of what the learner is using the language for. In general terms, he should be using the language 'in order to exchange thoughts, feelings and wishes with [others]' (Jespersen 1904:4); 'to make things happen' (Hodgson 1955:97); 'to use language to some purpose, to communicate and be communicated to, to assume certain roles' (Corder 1973:49); 'to make contact with each other as people, to exchange information and opinions, talk about experiences...' (van Ek 1975:ii). These are general aims, formulated for a variety of contexts at intervals over ninety years and imputed to learners by teachers and theorists.

A second strand to purposefulness is the commitment of self to the communication in progress: the 'thoughts, feelings, wishes' and 'information, opinions and experiences' necessarily imply the expression of some individual, personal predilections in the communication. If the classroom work is to be so organised that 'students feel the need to communicate in the L2' (Ellis 1990:126), if the student 'starts with something that he wants to say and with a person to whom he wants to say it' (Stevick 1976:107), the student must be given the opportunity to be himself, which leads on to the question of individualisation.

The literature on this topic over the past twenty years is extensive, indicating that it has been considered an important consequence of adopting a communicative approach to language learning. Individualisation is normally considered either in the context of providing self-access materials in study centres (Geddes & Sturtridge 1982, Sheerin 1989, 1991) or as ways in which students may be encouraged to relate personally to work in class, implying adaptation of method and material so as to engage the varying interests of the students (British Council 1978). This principle can be carried out of the classroom into the world outside, as with L1 helpers (Riley (ed) 1985) and the placing of students in job-like situations in the local community, such as teaching (Carter & Thomas 1986).

Two areas of discussion in language learning take an interest in the learner as an individual whose interests and capabilities are to be taken into account: language acquisition and learner differences. It might seem at first sight that they are interconnected, but the former is based on a theoretical construct and the latter investigates causes and effects. It is to be expected that, even if there is an underlying order of acquisition for all learners, as claimed by Krashen, each individual comes to the point of learning with a different set of expectations, experiences and abilities, so that the reaction to any input is different for each learner. But the rules are held to be universal, so that within Krashen's five hypotheses (the acquisition-learning distinction, the natural order hypothesis, the

monitor hypothesis, the input hypothesis, and the affective filter hypothesis - expounded in, for example, Krashen 1982) it is important that there should be escape routes. Language acquisition may follow a set order, but it is clear to those not entirely convinced by Krashen's arguments that individual differences will occur. The power of the monitor to tempt the learner into producing correct forms will vary with the individual; the affective filter will be silted up with varying innate deposits of motivation, self-confidence and anxiety. Both the monitor and the affective filter hamper the pure acquisition process and are presumably to be excused as inbuilt failings. Skehan (1989:3) points out that the monitor does not connect up with other aspects of the model, being the product of learning rather than part of the process of acquisition; and McLaughlin, in the course of a devastating attack on the whole of Krashen's approach, suggests that he 'has provided ... no basis for relating the affective filter to individual differences in language learning' (1987:55).

Learner differences in aptitude, motivation and language learning strategies are clearly related to the concept of individualisation in a communicative classroom, provided they can be put into effect. Skehan, after reviewing a wide range of research on individual differences in second language learning, concludes that in practice, the choice provided by individualised learning materials seems to have provided 'choice for its own sake, rather than principled choice linked to learner characteristics' (1989:140). Nearly ten years on, it seems that actual research studies are still not exploring the existing theories, and that 'there is scope now to explore just how instruction can be adapted to take account of the person who is most involved, the actual learner' (Skehan 1998a:281). The formation of an IATEFL Special Interest Group for Learner Independence in 1986 is an indication of the perceived importance of this concept for teachers, but their interest is apparently less in the possibility of principled variation in materials and teaching method to accommodate individual characteristics than in sociological aspects of learner autonomy, such as learning how to learn and self-assessment, as reported for example in Allan & Timmer (eds) (1996).

Principle 4

Error: Different approaches to error are required according to the kind of exercise the students are engaged in.

'Traditional syllabuses,' maintains Brumfit (1979:187) 'have always had a basis in the accurate construction of the target language, rather as if it were a building being built to a blueprint.' With a grammar-translation syllabus - a misnomer according to Howatt, for 'it draws attention to two of the less significant features of the approach' (Howatt 1984:131) - error is to be blundered into and constantly corrected (Harrison 1975). According to the behaviorist theory which underlay the structural approach, the student must get everything right from the beginning, on

the principle that early habits, bad or good, will be permanently retained. So learning by heart is important, repetition (eg through drills, facilitated by the use of language laboratories) is a means to an end and correction becomes endemic in language classrooms (Dakin 1973).

Corder (1973) attributes to Selinker the invention of the term 'interlanguage' to describe the unstable but systematic developing language of a learner. And it is Corder who defines most patiently the differences between one kind of 'error' and another. Native speakers have lapses (slips or false starts), and make errors (breaches of the code) and mistakes (producing inappropriate utterances such as social gaffes or wrong use of register). Learners also make errors (though these are the result of utterances in a 'different language' like that of a child learning L1), have lapses (though these are of no immediate relevance) and make mistakes (though these may be acceptable when they would not be with an L1 speaker) (Corder 1973:259-61). The object of error analysis is therefore 'to describe the nature of the learner's interlanguage and to compare this with the target language' (ibid:274).

But the starting point of this discussion was the need for varied attention to error, depending on the kind of activity being demanded of learners. The balance required between accuracy and fluency as requirements for learners' output has been a recurring theme in Brumfit's writing. A summarising statement of the issue is:

> However the syllabus is organised,... the teacher will have to offer some opportunities for work which is aimed at accuracy, and other opportunities for work aimed at fluency. Important as the accuracy work may be, it will be during the fluency sessions - whether silent reading, or group discussion of various kinds - that the strategies for independent learning, and indeed the real internalisation of the language, will be being developed.
>
> Brumfit 1981b:49

There are social as well as linguistic reasons why learners' errors should not be 'corrected' to conform with a notional native speaker norm. Widdowson points out, as Corder had earlier (1973:260), that poetry is deviant language. The violation of conventional norms is the result either of deficiency (= error) or of heightened proficiency (= art) (Widdowson 1984:246). He adds:

> So to correct learner language without communicative warrant in order to bring it into line with what an alien society regards as proper comportment is to impose standards of behaviour which in all probability the learner cannot identify with and in which he has no social investment.
>
> Widdowson 1984:249

The important phrase in this comment is 'without communicative warrant'.

Principle 5

Realism: The method operates with real language in real situations.

If, as maintained above by both Morrow and Johnson, language is to be used by

learners to communicate with a purpose outside the classroom, and with practical applications in view, realism is a factor which has to be taken into account. The first questions here are however, What is real language? and What are real situations? The situations come first, for they call forth the language.

The immediate reality is the classroom itself, and this is where Natural methods begin, as in Gouin's series, for example, with the students suiting the word to the action. "'I walk towards the door/I draw near to the door/... I get to the door/I stop at the door. I stretch out my arm/I take hold of the handle/I turn the handle...'" (Gouin 1893:131). This may look quaint to the modern eye, but it is a forerunner of the use of the classroom as a resource. Any other 'real' situations are in practice only representations of reality, for skills work is done mainly for teaching purposes: the text read or heard is only incidentally a means of cultural enlightenment; the writing and speaking are mainly done to practise writing and speaking. The nearest the learners approach to real situations for using language in the classroom is in activities such as role play (in which they are being someone else), simulations (in which they may be being themselves, but in fictional circumstances), group discussions (in which they may give their opinions, but perhaps on topics which do not essentially concern them), and finally in what might be called 'commitment' activities (in which they do give something of themselves, but to peers and a teacher rather than in self-motivated interaction with others). The opportunity for learners to use language in real situations may occur outside the classroom, in various contexts: for courses in Britain (or other L2 environments), communication with host families, where they occur, is inevitable (and desirable), and for some courses there are more deliberately introduced contexts such as sports tuition as complementary activities to language courses (see ARELS 1994) and arrangements for special project work in the local community (eg Carter & Thomas 1986). But the impact of these is not normally evaluated as an element in the learner's progress with language.

All these instances are aspects of 'authenticity', a concept which has provoked endless discussion from the earliest days of the communicative approach. Wilkins launches the debate with a definition:

> ...authentic language materials. By this is meant materials which have not been specially written or recorded for the foreign learner, but which were originally directed at a native-speaking audience. Such materials need not even be edited, in the sense that linguistically difficult sections would not be deleted, although the linguistic content of such texts could well be exploited in various ways.
>
> Wilkins 1976b:79

This is authenticity of text. But as Clark points out, there are other kinds of authenticity to be considered. Bringing into the classroom a wide variety of real bits of language (tickets, advertisements, travel brochures, timetables, menus...) will not ensure that what the learners do with them is necessarily useful (Clark 1987:149), let alone 'authentic'. Clark suggests that there are authenticities of

purpose, of response, of conditions, and of purpose to the learner as learner (ibid: 206). There is also the question of authenticity of task: are the learners being asked to do something with the material they have been given in a way which reflects their likely involvement in real life outside the classroom? Edge maintains that authenticity of task can be achieved by introducing an exercise which involves admittedly artificial division of information shared out to pairs, then to individuals, and applying it to a real-life situation (Edge 1984:259). Breen finds authenticity (though he does not use the word) in what each individual learner brings to a task - his interpretation of what is offered, which will be different for each one (Breen 1987:23-46). And yet another aspect of authenticity is what kind of output learners are expected to produce. Is it merely the appropriate response to a situation, or is it a matter of giving learners scope to express themselves in the foreign language? (Maley 1983:299). A common answer adopted by course book authors is to relate learning material to current events in a way which it is hoped will encourage the learners to respond, but then there is the problem of transience: the material cannot afford to be too closely related to the present or it will soon become unmotivatingly historical for subsequent learners.

It seems from the above discussion that 'real' language is easier for the teacher to provide than 'real' situations. The most important result of language learning must in the end be the ability to cope, no matter how this is achieved.

> We do not begin with authenticity; authenticity is what the learners should ultimately achieve: it represents their terminal behaviour.
>
> Widdowson 1979:166

The practical application of this dictum depends on finding material with which learners can use their available language and at the same time extend it in directions which they find authentic *for them.*

Principle 6

Choice: The students must have a choice of both ideas and language.

Choice implies a range of options, but at elementary levels the learner's options are by definition limited. Morrow explains that when the learners as speakers or listeners are controlled by the teacher, they are not learning how to speak realistically, that is, under the pressure that language use involves. Nor are they learning to understand when there is doubt as to what is to come next. They are not being given the chance to choose what they will say or to respond as they feel appropriate. Learners also need to anticipate if they are to make useful contributions from their own experience. Anticipation is necessary in all language use and language learning (Corder 1973:120), and this can be put into practice in the classroom, for example by setting tasks which allow for learner planning (Foster & Skehan 1996) and in reading exercises which ask learners to predict what a text may be about before reading it in detail (Grellet 1981:18).

Perhaps the common factor between these two apparently conflicting suggestions (spontaneity and anticipation) is the importance of appropriateness of utterance, so that in a communicative approach more time is spent on learning different ways of expressing the same function in different contexts, than on articulating the structures which underpin them. The question then becomes one of grading: how far teachers should go at any given point in the course in exposing learners to the variety which may be possible for the expression of any one move, for example the multifarious ways of asking the time as listed by Trim (1978:41-2). There is no doubt that learners do need a range of resources in terms of functional alternatives for picking their way successfully through the minefield of social solecisms: in the event, limits will be set by the contexts in which the learner will need to deploy his language skills in due course. For example, operating in the upper reaches of social life may demand less actual knowledge of the L2 than coping with day-to-day practicalities such as organising plumbing repairs. But the plumbing vocabulary may be found in a dictionary and illustrated by gesture, whereas failings in the inferences made and understood in the niceties of social interaction may have devastating consequences (Harrison 1980).

Morrow's proposition that learners should be given choice is taken further in his suggestions for using texts in a communicative approach. Choice is 'a key factor underlying involvement in a text', since in real life people choose what they will read or hear. He would like to offer learners not only a range of texts to choose from but also a choice of what to do with the chosen text (Morrow & Schocker 1987:252). Nunan (1989:20) also suggests that learners should be involved in designing or selecting tasks and allowed to choose what to do and how to do it. Even within a structured situation set up by the teacher for particular purposes, it is important that students should have the maximum opportunity to choose, deciding for instance what will take place, even making their own cue cards (George 1981:85).

> We need to provide consistently throughout the language-learning process occasions on which the learner expresses what he himself wants to express through the forms of language that are available to him at his particular stage of language learning.
>
> Wilkins 1974:76

The questions then are how to find out what the learner wants to express and how to accommodate the possible variety of wants within a single classroom.

Principle 7

Learning by doing: The students must be involved in what happens in the classroom.

Rogers (1983) identifies two kinds of learning. At one end of a continuum of meanings is 'the kind of task psychologists sometimes set for their subjects - the learning of nonsense syllables... Such learning involves the mind only'. But there is also significant, meaningful, experiential learning.

> *It has a quality of personal involvement* - the whole person in both feeling and cognitive aspects being in the learning event. It is *self-initiated*. Even when the impetus or stimulus comes from the outside, the sense of discovery, of reaching out, of grasping and comprehending, comes from within. *It is pervasive*. It makes a difference in the behavior, the attitudes, perhaps even the personality of the learner. (Rogers' italics)
>
> Rogers 1983:20

Applying these principles to language learning is a small step for communicative theorists. Wilkins characterises the situation with the comment that learning activities should be representative of learning objectives, which 'reflects that essential truth of the behaviourist view that "we learn what we do", while not insisting on too narrow an interpretation of *do*. We cannot learn what we have not experienced' (Wilkins 1974:59). On the other hand, 'learners are not necessarily best prepared for a given set of terminal behaviours by giving them practice in these behaviours alone' (Maley 1983:299). Being involved in what happens in the classroom implies a commitment of self not just to situations but also to developments and outcomes, which means joint responsibility with the teacher and other learners for the way in which activities work out and progress is achieved. Underhill shows how this can be done:

> The explicit belief of humanistic psychology is that under the right conditions of supportive non-interference, people can be self-directing and resourceful about what they need to do to make progress which they experience as significant.
>
> Underhill 1989:255

This not only sums up what 'humanism' offers language learning in terms of approach: it also encompasses the self-centred (in the best sense, 'self-aware' rather than 'egotistical') development of human relationships, in which L2 learning can play an important part.

Principle 8

Cooperation: The students should be working at tasks they cannot complete without the contribution of others.

The mechanical interpretation which may be put upon group activities such as 'jigsaw listening' (Geddes & Sturtridge 1979 - though not intended by the authors), regarding them merely as an arrangement of bodies in classrooms in formations which will engender language, falls short on the social implications. The point is that a properly communicative exercise cannot be completed unless every participant contributes his share. There is little benefit in a partly achieved end: success is either all or nothing. The same applies in principle through all language in use, for if no-one is listening, there is little point in speaking. Putting this into practice leads to an emphasis on group work, involving the sociological aspects of the classroom and the responsibility of the learners for shared progress. This would appear to be the converse of the need for individualisation as discussed earlier, but in a curious paradox, it is only by co-operating (in its most literal sense of operating

together) that the individual can learn how to cope with expressing his individuality to others. Even in competitive situations, the task cannot be completed without an opponent to confront and outdo: participants are cooperating even as they compete.

Eight principles

Any attempt at summarising the above commentary on the communicative principles proposed by Johnson and Morrow immediately comes up against the fact that they are not as distinct from each other as at first appeared. There is overlap in several areas, for example between application and realism, between information gap and purpose, and between learning by doing and cooperation. But the discussion has served to show how the communicative approach is claimed to be different from other theories of language learning. Nunan (1989:194-5) provides an overview of eight approaches and methods (derived from Richards & Rodgers), which differentiates theories of language, theories of learning, objectives, syllabus, activity types, learner roles, teacher roles and roles of materials. Nunan's chart is helpful in that it shows that there are many more concurrent views than might be expected about preferred ways to learn an L2, and that there is no clearcut differentiation to be made between communicative and all other methods.

Objections and continuities

Two recent papers which still, after 25 years, include 'communicative' in their titles (the concept is evidently still not worn out) are more critical than appreciative of the spread of the communicative approach. Ellis questions its universal relevance, offering evidence that in Far Eastern cultures considerable difficulty arises in the confusion between different world-views, and that the best the Western teacher can hope to achieve is that of cultural mediator. This means not just recognising the tradition of Chinese culture (and others similarly hierarchical) that the classroom is always to be teacher-centred, but discussing methodology with learners and showing empathy with the experiences of others (Ellis 1996:217). This is an exemplified extension of Maley's argument (1983), confirmed by experience in Japan (Sano et al 1984). Karavas-Doukas, in the other recent paper, writes that classroom studies (she refers to seven examples) seem to suggest that fully communicative classrooms are rare, in Greece at least.

> Broadly speaking, the communicative approach appears to have brought innovation more on the level of theory than on the level of teachers' actual classroom practices.
>
> Karavas-Doukas 1996

She goes on to report her findings from the use of an attitude scale with Greek teachers of English, which show that there was a considerable discrepancy between the teachers' classroom practices and their expressed attitudes. The question then arises of what criteria she uses to represent a communicative approach. They are derived from a direct practical example, since they refer to the Greek English

language curriculum and communicative textbooks introduced in 1987. They are: group/pair work; quantity and quality of error correction; role and contribution of learners in the learning process; role of the teacher in the classroom; place/importance of grammar. These are the course book manifestations of communicative principles in Greek classrooms.

Some attempt should now be made to sum up what communicative theories may mean as guiding principles for practice in language classrooms. It is difficult however to draw encapsulating conclusions from the elaborations of eight communicative principles as set out above. Any overall view is likely to refer to humanism and social context as guiding lights, for the emphasis is on offering the learner ways of achieving personal commitment in the learning of the L2, some choice in the subjects that the language will deal with, and continual reference to the people and situations influencing the appropriateness of the language used.

1.4 'Communicative' in practice

Course book interpretations

The next task is to consider how far the discussion of communicative principles over the last 25 years has been implemented in actual learning procedures. The simplest way of establishing the impact of this fresh approach is probably to explore what course books offer as suitable teaching and learning material. Eight 'post-revolution' course books have been investigated with this aim in mind, not necessarily those claiming specifically to be communicative, though all at least acknowledge the trend. So as to limit this choice and at the same time improve comparability, all examples are at Intermediate level. They are listed chronologically below, with brief indications of the reasons for inclusion:

- *Strategies* (Abbs et al 1975) - a pioneer and best seller
- *Quartet* (Grellet et al 1982) - deliberately espousing the new approach
- *Cambridge English Course* (Swan and Walter 1985) - best seller
- *Headway* (Soars & Soars 1986) - best seller
- *Cobuild* (Willis & Willis 1989) - based on a corpus of 'real' English
- *Workout* (Radley & Millichip 1993) - learner independence
- *New Headway* (Soars & Soars 1996) - best seller - what changes from *Headway*?
- *Lifelines* (Hutchinson 1997) - new - 'post-communicative'?

The most direct source of attitudes is authors' introductions, in which they lay out their offerings and explain their intentions. These expositions of the eight courses were compared in detail with the eight principles set out above, and the results are given in *Figure 1.1*. The notes in each of the 64 cells in the diagram are intended to represent the possible answers of the course book authors to questions about how far they put into practice the theoretical touchstones of the communicative.

Figure 1.1: Realisation of principles in course book introductions

COURSE BOOK

PRINCIPLE	*Strategies (1975)*	*Quartet (1982)*	*CEC (1985)*	*Headway (1986)*
application outside the classroom	• see how & where they can use the language they are learning	• with young adult students in mind	• adult students, for general practical or cultural purposes	• topics to appeal to mature adult not maturing adolescent
information gap don't already know	[not mentioned]	• problem-solving activity • resolutely task-centred	[not mentioned]	• speaking activities require information gap
purpose the individual	• help you to say what you want to say • what Ss need to express is most important criterion for selecting, grading	• not based on needs analysis • material which appeals to logical/rational & intuitive/creative • authentic reaction from S	• respect for the learner as an adult, with an individual learning style and adequate communication skills in L1	• Ss giving their real opinions, interacting as themselves • own opinions, experiences & feelings
error treatment	• communicate effectively rather than merely to produce grammatically correct forms • no hard and fast rule for corrections	• to favour acquisition & fluency • accuracy exercises in T's book only • sensitive monitoring • need for both grammatical accuracy & semantic appropriacy	• mistakes part of the learning process • exercises actually encourage fluency before accuracy • If communicate effectively not be given a sense of failure because of mistakes	• both accuracy & fluency • excess of attention to accuracy would stop the flow
realism language, situations	• increase Ss' communicative repertoire in social areas (eg functions) • listen to & read authentic spoken & written English	• authentic output by learner as well as authentic input by T • authentic or simulated-authentic texts	• some material is authentic • recordings of natural unscripted English	• receptive skill [with] authentic & semi-authentic material • real or realistic activities
choice ideas, language	• talk & write about yourself, your life, your ideas • open exercise (oral) means Ss are free to introduce own ideas and information	• need of Ss to be left elbow-room to exercise own initiative in learning	• Ss participate actively in choosing what language they will learn • different learning styles & personal interests	• many free speaking activities throughout • Ss assume considerable responsibility for their own learning • preferred learning styles
learning by doing experience, extension	• Ss' educational, social & personal development	• emphasis on work done by Ss.T as counsellor in the last resort	• model sentences of Ss' own construction	• Ss involved in learning process • Ss challenged cognitively • interaction for language acquisition & assimilation
cooperation input, turns	• practise language in pairs & groups • work in pairs simultaneously	• cooperation between pairs or groups • opportunity to correct each other • cooperation [for] more effective learning	• many exercises require students to work in pairs or groups	• encouraged to work closely with peers

COURSE BOOK

PRINCIPLE	*Cobuild (1989)*	*Workout (1993)*	*New Headway (1996)*	*Lifelines (1997)*
application outside the classroom	• communicative problems of outside the classroom	• everyday situations, functional practice material	• target language contextualised • both pre-communicative & genuinely communicative activities	• for adults & young adults • contexts relevant to the interest & lifestyle of modern young adults
information gap don't already know	• to solve a problem, exchange relevant information	[not mentioned]	[not mentioned]	[not mentioned]
purpose the individual	• English that is important for learners' needs	• need to be given the chance to revise structures, extend vocabulary • writing target to correspond to learners' needs	• activities personalized, Ss talk about themselves, learn about fellow Ss	• Ss reflect on their own aims & needs
error treatment	• tasks primarily concernedwith fluency • presenting report has clear focus on accuracy	• grammar check that students have grasped the rules...if errors, Ts should correct them	• practice from mechanical to semi-controlled to free • test your grammar at start of every unit	• exercises graded from controlled to free
realism language, situations	• rich input of real language • all from authentic sources, unchanged	• recordings are authentic, full range of accents	• all reading & listening texts have authentic source but many adapted • Post Script - functional, situational & survival skills are presented & practised	• wide range of authentic and adapted texts
choice ideas, language	• good deal of responsibility lies with the learners	• extra exercises, self-study for more advanced students • 'learner training' targets to develop learner independence	[not mentioned]	[not mentioned]
learning by doing experience, extension	• encouraging students to think about language for themselves	• Ss reflect on target structures & 'create' rules themselves • effective learning strategies • awareness raising, speaking before reading	• encourage good vocabulary learning habits • Ss to keep records of the words they come across	• emphasis on grammar discovery • range of reflection, practice & use
cooperation input, turns	• tasks usually done in pairs or groups	• [Ts given] suggested modes of organising Ss, pairs or groups, report back to class	[not mentioned]	[not mentioned]

The first point to be made is that there are only nine occasions where there is no response to a communicative principle as listed above. This seems at first sight to indicate that authors have been aware (or keen to show awareness of) these principles and have incorporated them in their courses. However, seven of the nine gaps occur with courses which have appeared in the last six years (since 1993), which seems to show either that the principles are so well established that they are taken for granted, or that they are now wearing thin in the face of experience of the kinds reported by Maley (1983) and Karavas-Doukas (1996). It is likely that *information gap* is now considered obvious, but *choice* and *cooperation* have apparently fallen into disrepute: the promoters of self-realisation such as Rogers, Stevick and particularly Moskowitz (1978) are out of favour. On the other hand, the newer courses give emphasis to the learner as an individual who should be encouraged to learn how to learn and to work out systems and rules for himself. There is a general return to grammar exercises, revision, checks and tests (in *New Headway* there is a 'Test your grammar' section at the start of every unit, and 'the amount and the level of the grammatical input has been increased' [in comparison with *Headway*]). Grammar is also the starting point for all the main units in *Lifelines*. And it is now called plain 'grammar' with no mention of the soft-pedalling 'structures'.

Reading *Figure 1.1* vertically, the differences between courses up to 1993 are as much in wording as in substance, though *Strategies* specifically addresses the student rather than the teacher; *Headway* emphasises the importance of the individual (*choice, learning by doing*); CEC is concerned with teacher attitudes (*error*) and student involvement (*choice*); *Cobuild* is perhaps surprisingly short on comments about *realism*, in view of its data base; and *Quartet* gives most space to its orientation to students (*purpose, cooperation*) and its guidance to teachers (*error*). *Workout* is concerned about mixed-ability classes and variations in teaching conditions, and so provides extra materials and encourages student independence; *New Headway* is somewhat thin in its relationships to the eight principles, sidelining functions and situations into a 'Post Script' section; *Lifelines* is thinner still, its introduction to the teacher's book consisting of two pages of list-like exposition (compared with eight pages of justification and a bibliography in *Quartet*) which concentrate on the three key features of the course: flexibility, interest and coherence.

Considered horizontally, the diagram indicates that there is more to be recorded about *error, realism* and *choice* than about *information gap* and *purpose*. The implications are that the first three areas are considered important aspects of the courses as practical implementations of communicative theory, and the latter two either so trivial that there is no need to mention them, or so obvious as to be taken for granted, though the likelihood is that they are subsumed under other headings, such as *applications* and *realism*.

Demands on course book writers

The course book introductions summarised in *Figure 1.1* give an overall view of the problems which course book writers face in meeting the demands of the market while at the same time attempting to put into practical form the current ideas about language learning. Few of the writers represented in the list above seem to have found the necessary balance between marketing and substance. Perhaps the most fundamental problem with a communicative approach is the relationship between notions and functions on the one hand and grammar and vocabulary on the other. A course book needs to suggest ways in which functions can be taught without merely supplying lists of alternative responses for given situations - the 'phrase-book approach' condemned by critics of communicative syllabuses as little better than the traditional practice of learning vocabulary by rote. Notions may be translated from the abstractions of a thesaurus ('Abstract relations, Space, Matter, Intellect', to quote the first three classes in Kirkpatrick's 1987 edition of *Roget*) into words concerned with length, width, thickness; or time, days, months; and so on, but the same difficulties arise in helping the learner to interiorise them without the mindless repetition of the grammar drill. The most obvious answer to these problems is to set the learning of language in contexts so as to clarify the range of possible applications, but only *New Headway* and *Lifelines* are recorded as including this as a principle. Perhaps it is too self-evident for the other authors to mention.

A second difficulty is introducing change to teachers who are happy with their current methods (and also to students expecting language education on conventional lines - 'where's the grammar?' they are heard to ask). They need to be convinced that the new methods will be more productive than the old. For these traditionalists, only good experience with communicative ideas such as group work and interactive tasks will be convincing enough. But trials are rarely offered to teachers: course books are more likely to be imposed from above, for various reasons. For example, educational authorities may need to show that they are keeping up with current trends; publishers constantly need new courses as the best sellers are spread abroad and eventually become over-exposed, especially since 'authentic' material fairly rapidly becomes out of date; private schools need new material to guard against the boredom of students and possible staleness of teachers. Course book writers' response is either to find new trends to exploit (eg learning to learn) or to provide a wide range of material with the hope of offering something for every teacher outlook and every learner interest.

The overall impulse of a communicative approach on course books has however been towards conformity. The eight courses considered above are all aware of the consequences of promoting interaction between learners in preparation for life outside the classroom: group work, the importance of individual input, and above all the development of the skill of speaking.

1.5 Testing in the classroom

The main interest of this thesis however goes beyond teaching/learning into assessment, specifically the relationship between the language as it is taught and learnt in the classroom and the content of classroom tests. How are teachers expected to link them? There are two likely sources of information which may help. The first is teachers' books on how to write tests. A wide range of these is available, from Lado's pioneering 'teacher's book on the construction and use of foreign language tests' (1961) to Bachman & Palmer's account of 'designing and developing useful language tests' (1996). Those which specifically mention the need for direct links between learning and assessment include Harris 1969, Valette 1977, Harrison 1983b, Carroll & Hall 1985, Heaton 1988, 1990 and Hughes 1989. But all these authors deal with tests as an add-on mechanism at various points in the course and are concerned with the priorities of test-making rather than the integration of classroom activities and assessment as two components of a single process - learning. Weir seems to be alone in appreciating the force of the connection:

> It would be useful if the criteria employed in the assessment of language production on tasks could be related in a principled way to the criteria for the teaching of a skill...
>
> Weir 1993:40

The other source of testing linked to learning could be the test material included in course books, where it might be expected that the assessments would relate specifically to the course which has gone before. When they are not designed as directly relevant to external examinations (and understandably therefore imitate in learning activities the forms of test appearing in those examinations), course books often include tests, sometimes in a separate test book for students. For example *Meanings into words* (Doff, Jones & Mitchell 1983) - 'a course in grammar and its use in communication' which tries to ensure that 'the relations between form, meaning and use are perceptible to the student at every stage'- includes class tests which are 'designed to check students' progress every four units'. The content of these tests does reflect the content of the relevant units, but the formats are largely traditional: sentence rewriting, asking questions, gap-filling, vocabulary and composition. Another course book which includes tests is *The Cambridge English course* (Swan & Walter 1985). It is claimed to be a 'multi-syllabus' course, 'designed for adult students who are learning English for general practical or cultural purposes' - a description which might seem to imply communicative assessment. But the tests are of structure, vocabulary, pronunciation, listening and writing, all according to well-tried formats: gap-filling with verb forms, missing words, reordering of jumbled words, and so on. There is no context longer than the isolated sentence except in the listening test and in writing, for example, a description. Finally, the *Collins Cobuild English course* (Fried Booth, Willis & Willis 1989), with its new emphasis on 'natural' English as available from a large corpus of several million running words, including spontaneous conversation, might be expected to break new ground in the

tests it provides. These again relate closely to the material in the course, and do include more convincingly realistic material, both in terms of language used and visual representation, and more context is given for the points tested. There are some more discourse-based formats such as matching, but there is still a high proportion of multiple-choice, true/false and gap-filling formats on isolated sentences rather than tasks requiring the integration of skills and an overall communicative purpose.

None of these three sets of course book tests shows any appreciation of what Howatt regards as the mainsprings of communication: they do not make clear who is being talked to, the value to the student of what is being talked about, or what the point of understanding and being understood might be. None of them gives any indication of criteria or content for the assessment of speaking. None of them considers the integration of skills which seems to be a necessary consequence of a communicative approach. None of them sees assessment as an integral part of classroom events instead of an added chore for both students and teacher at designated points in the course.

Course analysis and relevant assessment

If classroom assessment is to reflect learning in anything approaching a full sense, the first essential is to establish what the learning in question has consisted of. Some kind of analysis of learning activities is clearly necessary, and the most obvious place to start is course books, since they are easily available and static. The rationale of the present thesis is that if an analysis system can be developed to capture what a course book demands of a learner, this same system may later be extended or adapted to record whatever may happen in any classroom, including ad hoc developments initiated by both teacher and learners. It is intended that as a result of the analysis, an assessment can be generated for a specific context (a particular class of learners with a particular teacher in a particular place at a pertinent point in the course), but the system from which the assessment is derived may be universally applicable.

The thesis therefore aims to develop a process of analysis followed by a process of assessment. The process of analysis is to take the form of a record of what a course book demands of the learner so that this can used as the basis for assessment. Some starting point has to be found which will lead to ways of actually producing both analysis and assessment material, less theoretical than Bachman's components of language competence for testing (1990:87) and more practical than Munby's operational instrument for syllabus design (1978:154-189). Course book analysis is often undertaken during teacher training (eg McDonough & Shaw 1993), partly as a way of exemplifying variety in practice and partly to encourage a critical approach to materials. But this kind of analysis takes a teacher's view of what a course consists of and its relationship to method. It is not interested in the detail of what the learner does, but in the content of what is to be presented and in what order,

and how the teacher will put it into action. An assessment of the kind envisaged in this thesis is interested in what uses the learners can make of the language they now have available to them to deal with events which they can see as relevant to their own personal view of the world. The question of what 'communicative assessment' might mean is discussed later, but for the moment the principle may be stated that what is meant here by 'assessment process' is something which gives a learner factual information on what he personally has achieved with a task undertaken in cooperation with other learners.

The learner as protagonist

The learner is therefore the centre of interest for this thesis. The strategy adopted for the course book analysis has been to consider the course book almost as if it were an interlocutor communicating directly with the learner. A class consists of individual learners, each of whom is presented by the course book with a series of activities designed to promote his learning, but he needs to understand what is required of him in response to its demands. From this appreciation of the importance of the individual there has grown the principle that the analysis of learning material should be done from the point of view of a learner. Looking at an activity set up by the course book, the question to be asked by an analyser is: 'I am the learner: what do I have to do to fulfil the demands of this activity?' This dictates the form of the analysis, which in principle follows the order of necessary responses from the learner when faced with an activity presented to him by the course book.

The process of assessment is intended to provide learners with tasks which they can reasonably be expected to cope with as a result of the preceding work (since the assessment is ultimately derived from the course book). At the same time, it will indicate where they stand in relation to the world outside the classroom (since the tasks will demand learners' discussion and decision-making on potentially 'real life' events). The tasks contribute to an overall structure which demands the involvement of learners as themselves (not acting characters, as generally in role-plays), making decisions on practical issues which could affect them (not on hypothetical politico-moral issues, as generally in simulations).

The argument which supports these developments has been drawn from several fields, principally the continuing discussion about what 'communicative testing' in EFL may mean (eg Bachman 1990, Alderson & North 1991), but also relevant systems of testing in secondary schools in Britain (eg Page & Hewett 1987, Pennycuick & Murphy 1988), the recording of achievement both in school and beyond it (eg Burgess & Adams 1980, Broadfoot et al 1988), and procedures which guide adult language learners towards autonomy (eg Riley (ed) 1985, Dickinson 1987).

1.6 Definition of terms

'Terminological traps'

Rowntree offers a useful starting-point:

> ...the field of assessment is full of conceptual quagmires and terminological traps for the unwary or the short-sighted.
>
> Rowntree 1977:9

There does indeed seem to be some difficulty in the literature in differentiating between the terms 'test', 'examination' and 'assessment', quite apart from 'evaluation', which is coming to be used as an umbrella word to cover anything at all to do with considered judgement, in or beyond education. To start the discussion at this, the widest end of the topic, it is instructive to consider the very broad view of Rea-Dickins & Germaine:

> Evaluation has a different overall focus and several different purposes from student assessment. While evaluation may be seen as a 'means' analysis (it is intended to serve the learning process), student assessment has a much more limited perspective with a focus on the 'ends' of learning in terms of what the learner has achieved at particular points.
>
> Rea-Dickins & Germaine 1992:5

At the other end of the scale from general to particular, Hughes succinctly defines a language test as 'any structured attempt to measure language ability' (1989:4). Sitting at some point between these extremes, Pilliner (1968) discusses examinations and tests, and offers several possible bases for distinctions between them: time (an examination is a group of tests 'all administered to the same persons'); hierarchy (professors examine, teachers test); and subjective assessments (examinations) as against objective ones (tests). Pilliner prefers this last means of differentiation, but would probably admit that his choice is somewhat arbitrary and affected by convenience for the argument of his paper (Pilliner 1968:21-2). Bachman suggests that 'what distinguishes a test from other types of measurement is that it is designed to obtain a specific sample of behavior' (Bachman 1990:20-1). This implies that nothing which falls beyond the sample required is to be allowed, a restriction which seems to ignore any creative use of the language by the student, unless originality is to be defined as a behaviour and then sampled in some way. The breadth of view needed to encompass further considerations of this sort would alone justify the use of 'assessment' instead of 'test' in the present case. But Bachman also comments on the terms 'assessment' and 'appraisal':

> There appears to be no careful delineation of these two terms in the measurement literature, although they have been used in conjunction with impressionistic approaches to measurement ... and large-scale programs aimed at measuring a wide range of characteristics.
>
> Bachman 1990:50

There are however some helpful definitions from fields other than L2 learning. 'Appraisal' is used in a quite specific sense by Tough (1976, 1979) in her work in L1 with primary school children (aged 7-13) in England:

> We have used the word *appraisal* then, quite deliberately and see it as quite different from the concepts of *measurement* and *evaluation*. We are not advocating that teachers should try to place the child in any category, or rank him against other children. *Appraisal* in our view means building up a picture of a particular child, being able to recognize what he is already able to do in using language and discovering what he may not yet be able to do by talking with him.
> (Tough's italics)
>
> Tough 1976:34

This approach is much closer to that of the present work, but 'building up a picture' involves for Tough taking time (up to half a day on each occasion) to find out what the child can do, and this is not possible in the EFL context, where the constraints of contact time and student numbers in classrooms have to be taken into account.

The Task Group on Assessment and Testing (TGAT), set up in 1986 'to advise on the practical considerations governing assessment within the national curriculum' (in England and Wales) defined 'assessment' as 'a general term enhancing [sic] all methods customarily used to appraise performance of an individual pupil or a group (DES 1988:preface). The report uses the term 'test' in the sense of 'any of a broad range of assessment instruments with standardised rules of administration and marking which ensure comparability of results'. These definitions seem to imply that 'test' equals 'assessment' with comparability attached. For Gipps, writing more generally in the context of assessment in education in the 1990s, 'assessment' includes tests, examinations, practicals, coursework, teacher observations and teacher assessment (Gipps 1994:2-3).

Discussing assessment in the wider area of education in general, Rowntree can afford a more sociological approach which at the same time appreciates the importance of the individual:

> ...assessment in education can be thought of as occurring whenever one person, in some kind of interaction, direct or indirect, with another, is conscious of obtaining and interpreting information about the knowledge and understanding, or abilities and attitudes of that other person.
>
> Rowntree 1977:4

This, as Rowntree later points out, includes both assessment by one learner of another and self-assessment by the learner, as well as the more traditional judgements of teachers about learners.

In sum, the term 'test' seems to be applied to mechanisms which are structured, objective, specific and in some contexts of relatively low prestige; 'appraisal' implies more general judgements, often spread over time and/or including a range of procedures designed to obtain information about various aspects of the student's standing; an 'examination' is regarded as a formal occasion which presents

candidates with various kinds of test; 'evaluation' is an all-embracing concept which includes judgements about administrative efficiency and value for effort spent - a market-forces approach; and finally, 'assessment' is either regarded as an umbrella term for all procedures which obtain information about student progress and status, or is used in a particular sense for a universal phenomenon - the way in which people respond to situations and to each other. As further evidence of how confusing attempts at definitions in this field can be, Rea-Dickins & Germaine (1992:3-4) use as an example of 'evaluation' the same sort of 'natural activity' among humankind as that suggested by Rowntree to illustrate 'assessment'.

Examinations vs assessment

Tests and examinations (the latter considered as formalised groups of the former) have always been regarded as having a considerable contribution to make to education. They may be welcome as a goal for teachers as well as for students, as a system which is more admissible socially than nepotism, and as an external and therefore impersonal assessment which lifts from the teacher the responsibility for decisions which could blight his relationship with student and parent (see Hutchinson & Young 1962, Montgomery 1965, Bruce 1969). But it may not be entirely unfair to attach to them concepts such as rigidity, unrealistic expectations of accuracy, undue influence on the curriculum, inadequate representation of learners' achievements, resistance to change and the dominance of administrative demands over learning priorities (see, for example, Pearce 1972, Burgess & Adams 1980, Broadfoot (ed) 1984, Gipps 1994). And there are wider implications, relating to social hierarchies:

> We try to discern merit in whatever form, particularly
> In the form we prescribe, the form of order;
> And so it appears
> Our young friend here
> May have to wait,
>
> Or else acquire the things we have in mind that all
> Right-minded persons sincerely approve,
> Viz., qualifications
> And subservience.
> Our time is short.
>
> Champkin 1965:26

Examination thinking and procedures seem unsuitable for present purposes. In contrast, 'Assessment' has associations with understanding, co-operation, positive attitudes, ipsative procedures ('in which the pupil evaluates his/her performance against his/her previous performance' - Gipps 1994:vii), attention to the individual, informality, continual review and 'freedom to learn' (one goal of which is 'helping students to prize themselves, to build their confidence and self-esteem' - Rogers 1983:3). Current practice in many classrooms incorporates 'humanistic' approaches (Stevick 1980, 1982, British Council 1982, Underhill 1989), in spite of some backlash against them (eg Gadd 1998 and Arnold 1998), and it would be interesting to

consider, as a means of further extending the concept, how far their principles can be applied in the development of assessment materials. Whether the assessment procedures eventually proposed by the present work do help to 'build learners' confidence and self-esteem' is perhaps for the learners themselves to judge. In any event, further investigation on these lines goes beyond the scope of the present work.

The justification for the use of the word 'assessment' in the argument of this thesis is that it includes much more than 'test', is not to be equated with the limited formality of 'examination', cannot afford the time needed for 'appraisal' and is less subject to market forces than 'evaluation'. So the current definition is:

> *For present purposes, assessment is a procedure for providing a learner with helpful information about his current status in specified aspects of his learning.*

Proficiency, achievement and diagnostic feedback

Several kinds of information can be provided by assessment which is helpful to learners, and also to their teachers. This thesis is concerned with two: information about the learner's application to realistic ends of the language which has been covered in the learning programme so far; and information about the components which have (and have not) been sufficiently absorbed into the learner's available language. The first of these requires the assessment of achievement (and to a certain extent proficiency) whereas the latter requires an emphasis on the diagnosis of adequacy (and inadequacy). Achievement relates to what has been learnt; proficiency to what the learner can do with the learnt material. Achievement looks back; proficiency looks forward. Both are interested in a general capacity - the learner's ability to cope with situations, to understand and be understood, to operate as a language user in the context provided by the assessment. The crucial differences between achievement and proficiency lie in the aims of the assessment and the source of its criteria for success.

Proficiency is defined and debated in terms of its component abilities, as set out in descriptive frameworks (eg Bachman 1990:42). These components are what Bachman and others call 'facets'. The existence of these components is postulated by speculative analysis of language in action (resulting in a patterned array of competences - see Bachman 1990:87), and the use of tests to test that individuals do use them in the appropriate conditions. Statistical treatment of the results of these tests provides the justification for attaching descriptive labels to the facets of the concept 'proficiency'. Tests designed to assess a learner's skills in the named facets will then show how far he is proficient. There is no interest in how this proficiency has been arrived at, only a calibration of its existence in a given sample of a learner's language use.

Achievement on the other hand has as its aim the recognition of progress in relation to preceding learning, and its criteria are therefore drawn from what has gone

before. Success is considered in terms of the learner's familiarity with the language he has met, whether in class or beyond it, and is reported back as a reflection of what he can do with it.

These descriptions have been presented as a dichotomy, but the relationship between proficiency and achievement is more complex. There are elements of achievement in proficiency, for the candidate who has had no class experience is rare; there is even more overlap with proficiency in an achievement test if it claims to show how the learner uses language for a purpose beyond the classroom, though this is the chief aim of 'communicative' teaching. In an attempt to clarify this potential confusion, a further definition is to be added to those suggested above.

> *For present purposes, the assessment of language achievement is defined as the process of reporting what language a learner has available from preceding learning to deal with situations representing his likely use of language in the future.*

In comparison, the definition of diagnostic assessment is relatively simple.

> *For present purposes, diagnostic assessment is defined as the process of reporting back to the learner and teacher with a detailed account of how the learner has used the mechanics of language, so that they may appreciate progress and recognise areas of shortfall for future improvement.*

The present concerns are therefore with quite restricted, though frequently needed, types of language assessment. Fuller discussion of the more general purposes of testing, as relevant to this thesis, appears in Whitehead 1962, Ingram 1977, Rowntree 1977, Page & Hewett 1987, Skehan 1988.

1.7 Method

The term 'empiricism' is applied to both a philosophical position and a method of investigation, but it is as a method that it guides this thesis.

> ... empiricism advocates the collection and evaluation of data. In this sense the focus is on experimentation and an empirical investigation is one guided primarily by induction from observation rather than by deduction from theoretical constructs.
>
> Reber 1985:238

In terms of assessment, this translates into a concern with content validity rather than construct validity, since what is of interest is the data appearing from the observation of relevant material rather than data resulting from the investigation of theoretical models.

In brief, the aim is to produce from the summarised results of an analysis of teaching material (input as found) a set of assessment materials which are related specifically to the learners' current standing (repertoire as found), and which require them to use the language they now have available to them (output as found) to solve by

discussion and decision some living problem (context as found). Throughout all this, the researcher's decisions are inevitably modified by a variety of experiences - of course books, of teaching situations, of teacher and learner reactions. But there is one overriding obstacle to objectivity:

> We recognize that it is never possible to be completely objective, that all experiment is contaminated by the presence of the observer, not so much because he is there but because his results depend on his view of what he thinks he sees.
>
> Davies 1977a:3

1.8 Preview

In the concluding section of a paper commenting on communicative testing in the previous ten years, there appear some suggestions for the future, among which are several which are taken up in this thesis:

> ...how to control what the test demands of the student without pinning him into a situation where he cannot show what he could be capable of; how to bring a dynamism into assessed exchanges which will go beyond the mundane buying of the railway ticket, and so on. The encouraging moves of the moment are those which investigate what the student is being asked to do, what the text consists of, looking for the construction of communication as well as its constructs.
>
> Harrison 1991:104

The following chapters attempt to put this theory into practice. Chapter 2 offers an account of what 'communicative' may mean for the assessment of learners' available language (a phrase used henceforth to describe whatever language a learner can bring to bear on a given task at a given moment). Chapter 3 takes up and investigates key areas of likely difficulty in a research project intending to devise language assessment systems which reflect a communicative approach to learning. Accounts of three stages of such a project then follow: Chapter 4 describes the development and use of a system of analysis for course books; Chapter 5 deals with the derivation of assessment specifications from the analysis; and Chapter 6 reports on the development of assessment materials and a series of trials with various groups of learners. Chapter 7 then discusses systems for making judgements about the spoken texts which these learners created and how useful feedback may be presented to learners and teachers. Finally, Chapter 8 reviews the work done.

Chapter 2

Assessment and communication

2.1 A sequence of interpretations

Introduction

Chapter 1 has included a discussion of communicative language teaching and learning as a first step in identifying classroom activities, with the aim of eventually setting up relevant classroom assessment. This chapter now considers developments in applying the principles of a communicative approach to testing and assessment, both in theory and in practice. It starts with a historical survey which looks for early signs of communicative ideas in what are usually considered to be traditionally-inclined texts on testing, brings the discussion forward to include communicative theory in testing, and then goes on to explore how far this theory has been reflected in practice in existing tests and examination systems.

Form and use

Concern with the application of language in use rather than solely with its form appears, rather surprisingly in view of his reputation as the diehard structuralist (see Davies 1978), in Lado's classic teacher's book on 'the construction and use of foreign language tests' (Lado 1961). His chapter headings may indicate an atomistic approach - eg sound segments, grammatical structure, vocabulary - but he asserts that the aim should be to test 'language in use' rather than the training which went into the learning (46). He discusses the view of speaking as 'the ability to express oneself in life situations, or the ability to report acts or situations in precise words, or the ability to converse, or to express a sequence of ideas fluently' and quotes a 3-point rating scale 'describing the responses to be expected and the score to assign to each' (240). But he is sceptical of the scale, preferring an approach through itemised 'signaling systems', in which 'the situations become of secondary importance, since they must not be a decisive factor in the test' (241). His overall thesis is that it is the difficulties for the L2 learner which need attention in testing, with the rest being ignored (29, 246), and this has parallels in principle with current ideas about the assessment of language in situations, where the context indicates what elements are critical for communication, and other aspects such as formal accuracy may be considered less important. But Lado does not allow context as a testable element, in contrast to the present insistence that context affects the appropriateness of the language used. In both instances, there are critical points to be included in the assessment: the interest in precision is now concentrated on

what language is right for the situation rather than on Lado's concern about what language is 'right' in a more normative linguistic sense. The nub of the communicative assessment problem is that precision is not a quality inherent in situations, since variability and fluidity are essential aspects of the individual's reactions to his surroundings.

Lado's argument is that situations are not a sufficient basis for generating achievement scores because they are too varied to be sampled and in any case do not necessarily require knowledge of language for successful participation. 'The elements of language on the other hand are limited, and it is more profitable to sample these elements than to sample the great variety of situations in which language can be used.'(27) Consistently with Lado's guiding philosophy, the reading and listening texts contain problems (in this case for speakers of Spanish) and the students' comprehension is tested 'precisely at those points at which the problems are crucial' (232). In speaking tests, again, 'the general technique is simply to give the student sufficient clues to produce certain utterances that contain the problems we wish to test.'(242) But Lado then goes on to offer two sample topics for role plays, one describing to a Spanish-speaking person 'the town or city in which you live', the other trying to persuade a German to emigrate to the United States (244). It is difficult to suggest what specific points will occur inevitably, or even regularly, in these speech samples so that the kind of testing Lado demands can be undertaken. On the other hand, they do not relate directly to the principles of purpose and realism as characterised in Chapter 1: for example, the candidate has no particular person to address and so has no indication of what interests to respond to.

In due course, the reputation of Lado's book inevitably fades until it is regarded as only 'a pioneering work which retains its value as a comprehensive guide and a check list of test types' (Davies 1977b:104). It has become apparent that restricting testing to points of difficulty not only implies a different syllabus for each pairing of L1 vs L2, but cannot cover enough ground for overall strengths and weaknesses to be reported. But Lado's definitions still hold as useful reference points, for example his differentiation between achievement and proficiency.

> The tests we have discussed in this volume attempt to measure how much of a foreign language a student knows. Such tests are usually called *achievement* tests, making reference to the fact that students have to struggle through a course or a learning experience of some sort to "achieve" a certain amount of control of the language. When the same tests are thought of independently of the learning experience, they may then be referred to as *proficiency* tests. Proficiency tests measure how much of a foreign language a person (not necessarily a student) knows.' (Lado's italics)
>
> Lado 1961:369

This is an admirably clear benchmark against which to consider later imprecisions. A significant problem for the present work lies in which of the two labels should be applied to the kind of classroom assessment which is to be developed: achievement because there is a direct relationship between the assessment and the previous

learning; or proficiency because the aim is to show how far the learner may be able to use this previous learning to operate in the real world beyond the classroom. Potential confusion between the two terms for this work stems from the distinctions to be made between competence and performance. The contrast has been made from different points of view depending on the interests of those making it. The clearest instance of this differentiation is between competence as the underlying inborn facility with language, as contrasted with performance as the outward manifestation of it - what language users actually do. This view is elaborated in Chomsky's writings (see Allen & Buren 1971) and pursued by Hymes (1972) into social applications. The division is at base a philosophical one, but comes to be used as a contrast in terms of testing and assessment, as a practical issue (Morrow (1977) and as a theory relating competence to performance (Canale & Swain 1980, Canale 1983, Bachman 1990, McNamara 1995 and Skehan 1998a).

The issue is important here because it affects both specifications (content) and judging systems (marking). If the main issue is how well the learner has mastered the syllabus (representing course book content and any other additional objectives proposed by the teacher or by various interests external to the classroom), the assessment needs to show only that the learner can respond to exercises in language use which represent what has been offered for learning. If on the other hand the main issue is how well the learner uses the language he has available (which he has gathered from all useful sources, both in the classroom and beyond it), the assessment needs to show how far he can apply himself and his skills in tasks representing real-life applications. The requirement is then not only knowledge/skills but also commitment, since without it, the interaction required by the tasks cannot be successfully pursued to a conclusion. These suggestions are for the moment speculative and will need to be elaborated, so from this point onwards, reference will be made to various thinkers' conceptions of 'achievement' and 'proficiency' in an attempt to build up a constructive definition for present purposes.

A different approach to language testing is introduced by Carroll, who discusses 'the psychological principles underlying the elicitation of language behavior and the interpretation of that behavior' (1968:46). The phrasing of this statement makes clear that Carroll is still working within the 'behavioral' tradition, Spolsky's (1978) 'psychometric-structuralist' phase, but he breaks new ground in proposing that it may be enough from a practical point of view 'to construct tests that measure only integrated performance based on competence' (Carroll 1968:56). He then adds an important rider about intentions.

> Apparently the extent to which a language test should attempt to measure specific aspects of competence depends upon its purpose - that is, the extent to which there is need for diagnoses of specific skills as opposed to a generalized, overall assessment of proficiency.
>
> Carroll 1968:57-8

Davies (1977b) approaches the problem of coverage by considering 'a new kind of examination' which attempts to build in both psychometric objectivity and linguistic realism and to work with detailed reference to a syllabus (52). This new approach, he reports, has looked at the needs of the pupil and so contains 80% speaking/listening rather than 80% reading/writing. (The needs are hypothesised, not the subject of analysis, as envisaged by Richterich & Chancerel 1978). Davies builds up a comprehensive syllabus from a matrix of components, plotting skills against levels. Levels in this instance are not grades of difficulty but categories of language, what Davies calls 'levels of linguistic analysis' (66), from Phonology, Grammar and Lexis to Context and Extra-linguistic, a categorisation influenced perhaps by Halliday's concept of 'delicacy' (Halliday et al 1964). Davies develops this suggestion in a series of matrices which progressively illustrate his panoptic coverage of testing, each cell in the final overall matrix containing an example of possible test content (72). He then gives examples of test techniques which fulfil the requirements for each of the 20 cells (79-109).

The most likely areas for finding the seeds of communicative testing, if any, in Davies' matrix are the levels slots labelled *Context* and *Extra-linguistic*. For example *Speaking: Context* has a question-answer test in which the candidate is asked to give 'what seems to you an appropriate answer' and a picture description which becomes a story; *Speaking: Extra-linguistic* is the equivalent of the free essay and takes the form of completion of a picture story. But these production tests, having rather reluctantly envisaged simulated real conditions, then treat them in a traditional way by assessing the language used rather than the meanings achieved.

In Davies' survey article on language testing (1978) discusses several issues raised by the differences between the two editions of Valette's *Modern language testing* (1967 and 1977). He points out that though Valette quotes Spolsky's (1978) three stages in the development of language testing - prescientific, psychometric-structuralist and psycholinguistic-sociolinguistic - she gives examples of only the second. He suggests that Valette's lack of examples for the third stage may be because of lack of rigour in these forms of test (149). In fact, however, Valette discusses all the 'prescientific' tests: translation both into (260) and from (209) the L2, composition (11 references) and items of grammatical, textual, or cultural interest (passim). She claims moreover that in the latter sections of her chapter on Methods of Evaluation,

> the emphasis shifts [from discrete-point items] to the testing of communicative competence and suggestions about other global tests of language ability, such as dictations and cloze procedures, which measure the students' ability to function when there is reduced redundancy (that is, when elements are missing or difficult to understand).
>
> Valette 1977:310

Valette's view of the achievement/proficiency distinction starts from the standard comment that achievement measures how much the student has learned in the course of instruction, but makes the point that not all programs are the same (5): the implication is that achievement tests are representative, and sampling therefore a crucial issue. The aim of the proficiency test is to determine whether the language ability measured corresponds to specific language requirements. This means that she regards proficiency as cutting across placement (ready for a course?), achievement (capable of reading professional literature?) and prognosis (skilled enough for a Foreign Service posting?), even though from a different perspective all these situations could be regarded as representing achievement of one kind or another. However, one of her comments could have been made as the first step in an argument for the change to a communicative approach in testing:

> The aim of the language course is the development of communication skills. Communication is a meaningful activity, and the test items should, as much as possible, be presented in a meaningful situational context.
>
> Valette 1977:4

The equivalent passage in Valette's first edition of 1967 speaks only of enjoying literature, appreciating the culture and conversing freely (4). It must be admitted that Valette appears to have caught on to the new idiom but not worked out how fundamentally it may affect testing practice, and this is perhaps the point which Davies was making in a different way with his comment on the lack of examples.

The notion of integrative testing is developed by Oller into a wide-ranging principle, by which he maintains that there is common ground to all language tests, as represented in cloze and dictation. These and other 'pragmatic' tests could be substituted for more traditional tests, saving the time and detailed attention required by, for example, Lado's separate tests of elements contributing to language proficiency. Oller's approach to substantiating this idea is through correlation studies which he claims (Oller & Streiff 1975, Oller 1978, Oller 1979:58) answer the problems presented by the testing of communication by focusing on candidates' 'expectancy grammar'.

Oller's concern with normal constraints and pragmatic mappings leads him into claiming that there is only one competence which needs to be assessed. This Unitary Competence Hypothesis (UCH) is parallel to another contemporary concept, General Language Proficiency (GLP): the similarities and differences are discussed by Vollmer (1981), but he finds neither convincing in the light of the research evidence (166). Contributions to the discussion by Hughes (1981) and Davies (1981) explain their scepticism about the statistical methods used. The debate continued for some years, with both empirical research studies and discussion of the qualities of tests and subjects. These developments are recounted by Skehan (1984, 1988), who summarises: 'the statement of the UCH generated an enormous quantity of research and increased understanding of testing and analytic techniques. But... the extreme form of the UCH is now untenable' (1988:213). Some of Oller's

suggestions for 'pragmatic' tests live on, chiefly dictation and cloze tests, but the fact is that they cannot fulfil the claims he makes for them.

A further development of testing by gap-filling is cloze is the C-Test, which is claimed to be an advance on cloze, with its 'rather considerable problems' (Klein-Braley & Raatz 1984:135) This technique is further elaborated by Klein-Braley (1985). Instead of deleting words, the C-Test deletes the second half of every second word. But the C-Test has its own problems, for example reading difficulty and the fact that responses may be found in the surrounding text (Hughes 1989, Jafarpur 1995).

A useful account of the background to cloze, C-Test and dictation is offered by Hughes (1989), who gives explanations of various kinds of cloze, with examples, then a rationale for the C-Test and an example, then a description of dictation and how it may best be used. But from a communicative point of view there are still fundamental problems with all three: none of them allows spontaneous production by the candidate; there is no oral production; and all depend on a knowledge of the language system rather than the ability to operate it (Morrow 1977).

Oller's approach to achievement as opposed to proficiency is in effect to deny its separate existence as a concept. But he admits it is a major concern in schools, where the curriculum is commonly tested at intervals, and especially at the end, with what are normally known as achievement tests, relating specifically to what has been covered in the course.

With Oller's hypothesis falling into disfavour, alternatives needed to be considered, if indeed they had ever been discounted by the mainstream of testing specialists in view of Oller's theories. If not one, several; if not unitary, plural. Carroll's (1968) approach again becomes relevant: he looks forward past Oller to the argument that context includes situation, for it is only a small step from pointing out the importance of textual embedding in the testing of language to considering where and when the text occurs. (This applies both to texts presented to the learner and to those produced by him.) The importance of context is emphasised by Rea (1978), who suggests that this includes linguistic creativity, defined as the ability to generate 'novel' and appropriate utterances, and the ability to communicate within a specified situation, both of which are best measured directly by genuinely communicative settings.

These steps towards the consideration of context are encouraged by the writings of sociolinguists (eg Pride & Holmes (ed) 1972, Giglioli (ed) 1972) and become a key consideration in commentaries on the communicative approach to learning (as reported in Chapter 1). Hymes' (1972) definitions of competence and performance are especially relevant when the discussion turns to testing, for a vital consideration is whether it is the learner's capacities or his applications which are the subject of assessment and hence of the information gathered from him and about him and, in the final analysis, for him.

Spolsky's often quoted history of the development of language testing in three phases (1978) - prescientific, psycho-structural and integrative-sociological - places communicative language testing in the last phase. For him, the main new departures are firstly an emphasis on 'the creative element of language, the infinite nature of the set of possible sentences, and the incompleteness of grammars attempting to characterize knowledge of a language' (ix). Secondly, there is now a 'need to test communicative competence, which includes the notion that knowing a language involves being able to use it in certain circumstances...[and]...allowing for the knowledge of different varieties and the ability to handle them in different circumstances' (ix). His approach at this time is thus shaped both by the aftermath of Chomsky's theories and by the impact of Hymes and others on the sociological front, with the balance in favour of competence rather than performance as the criterion for validity.

What the introduction of sociological concerns into the language discussion means for testing is explored by Morrow (1977, 1979), who suggests that features of language use which are not measured in conventional tests must be taken into account if a test is to be considered communicative. These are:

- interaction - 'What is said by a speaker crucially depends on what is said to him';
- unpredictability - the problem of processing language in real time;
- context - both situation (physical environment, participants' characteristics) and language (textual coherence);
- purpose - why things are said and how to achieve one's own ends;
- performance - accepting the less-than-ideal limitations of spontaneous use of language and strategies for coping;
- authenticity - mainly simplification of text; and
- behaviour - 'A test of communication must take as its starting point the measurement of what the candidate can actually achieve through language' (1979:150)

Morrow then goes on to consider what these criteria imply for tests as mechanisms for producing information about students. A test of communicative ability will have at least the following characteristics: it will be criterion-referenced; it will be concerned with content, construct and predictive (but not concurrent) validity; it will rely on qualitative rather than quantitative modes of assessment; and it will be more concerned with validity than with reliability: 'Spurious objectivity will no longer be a prime consideration' (p151). He argues that concern with performance is quite novel for tests and points out several problems with it, for example generalisability - whether performance in one situation can be taken to represent

an equivalent performance in another - and the ways in which qualitative assessment is to be implemented.

Following on from this argument, Morrow suggests that performance tests are necessarily integrated (the whole of language in use being more than the sum of the parts) and that for proficiency testing, 'it seems incontrovertible that performance tests are necessary'(1979:151). For diagnostic testing, on the other hand, he considers that discrete-point tests might be useful, as exemplified in the 1977 paper.

Morrow's view of achievement and proficiency is that performance (ie integrated) tests should be used for both purposes. Proficiency tests will be offered in terms of specified communicative criteria, which means: that pass/fail gives way to reporting what a candidate can do; that profile reporting (on separate skills) is helpful for differentiated assessment; and that syllabuses will need to specify types of operation, content areas and criteria for assessment. But achievement testing should also consist of performance tests because Morrow sees it as 'misguided' to use discrete feature tests for this purpose.

All this chop-logic about the terms Morrow uses may seem finical, yet it is important that defensible labels are attached to the purpose and scope of an assessment system, and in particular any system which may be proposed for a designedly learner-based specification, as in the present case. If the learner is to work at tasks to demonstrate his ability to use the language he has learnt so far, it is important that the purposes of the tasks, and hence the purposes to which he is applying his language skills, should be clear not only to the writer of the materials but also to the learner involved, so that he can see the relevance of what he is doing to his life beyond the classroom.

These two publications of Morrow's (1977 and 1979) were prepared as part of the development of a new communicative examination for the Royal Society of Arts (RSA). In a contemporaneous but separate project, the Council of Europe were concerned that the Threshold level specification for a communicative approach (van Ek, 1975) should be represented in a test. A series of experimental tests were prepared and trialled, culminating in one example of a T-level test, developed over three years in five countries (Groot & Harrison 1979). This was mainly traditional in format, and a paper was commissioned to explore more radical possibilities. The purpose of the paper was to discuss the assessment of 'the "application of T-level" rather than the "acquisition of T-level"', in other words, taking a proficiency rather than achievement perspective on the problem (Harrison 1979:1). The discussion is summed up under 'evaluation rules' for each criterion: realism; redundancy; the necessary response; appropriate use; natural use; practicality; non-linguistic factors. These abstract nouns are supported by examples of linguistic events and potential circumstances, but the discussion now seems somewhat repetitious and yet at the same time incomplete. Nevertheless, the ideas remain a continuing undercurrent to the present work.

A theoretical set of principles for suggesting what communicative learning (and following on from it, communicative testing) might consist of is Canale & Swain's framework of constructs (1980). They make several important contributions to the 'communicative' debate, in a paper which has been frequently cited as a source of inspiration for later writers in the field. The first contribution is a detailed discussion of the relationships between communicative competence and communicative performance. They conclude that the former is part of a more general language competence and the latter part of a more general language performance (1980:7), but make no attempt at succinct definitions of the two concepts. They do, however, set out five principles which they consider essential for the development of a syllabus adopting a communicative approach (and since their concern is with testing as well as learning, it may be assumed that the same principles apply in a communicative approach to testing):

- it must aim to facilitate the integration of three competencies: grammatical, sociolinguistic and strategic;
- it must take account of learners' communicative needs, both fixed and interim, in these three areas;
- it must allow the learner the opportunity to respond to genuine communicative needs: 'We think that exposure to realistic communication situations is crucial if communicative competence is to lead to communicative confidence' (p2);
- it should make use of those aspects of communicative competence acquired with the first language;
- it should provide information, practice and much of the experience needed to meet learners' communicative needs in the second language, and also build on some awareness about language and culture (Canale & Swain 1980:27-8).

The notion of 'communicative needs' recurs frequently in the elaboration of these five principles, but the phrase is evidently interpreted in a vaguer sense than that envisaged by Richterich & Chancerel's (1978) identification of information from various sources to help the learner make decisions, and even further from the comprehensively specific requirements listed by Munby (1978) for designing syllabuses. In sum, Canale & Swain associate a communicative approach with more realism for learners so as to meet their 'communication needs', without defining what 'communication' implies for methods or materials, or indeed what 'needs' might entail in this context.

Another of Canale & Swain's contributions is to undertake a discursive survey of a wide range of opinion which leads them to propose a theoretical framework for communicative competence which rests on three main competencies: grammatical competence, sociolinguistic competence and strategic competence. (Canale later

(1983) added a fourth component, discourse competence, the ability to handle language beyond sentence level.) The label 'grammatical' includes not just grammar as understood in texts used in traditional language learning, which Spolsky would call 'prescientific', but also lexis and phonology - ie the total of the mechanical aspects of a language. This leaves 'sociolinguistic' to cover rules of use and rules of discourse and 'strategic' as the means of coping with breakdowns in communication, either grammatical or sociolinguistic. Within each of their three components Canale & Swain propose probability rules of occurrence, a suggestion which leads on to consideration of authentic texts and of the availability of choice to the learner in his language production (31).

The first implication for testing is that both competence and performance must be tested (34), even though it is 'important to keep in mind that one cannot directly measure competence: only performance is observable' (6); but at the same time 'communicative competence...is observable indirectly in actual communicative performance' (29). The second implication is that both discrete-point and integrative tests should be used in their proposed communicative approach because the former may be more effective in making the learner aware of, and in assessing the learner's control of, the separate components and elements of communicative competence (34-5). This sounds more relevant for what is usually termed 'diagnostic' testing, in which the aim is to give detailed feedback on points which the learner might with profit review.

Terminology is straightforward for Johns Lewis (1981): she adopts a definition of communicative competence from Hymes, meaning something more than a functional approach, and including all the non-linguistic factors beyond the spoken/written word which contribute to the social meaning of an interaction. She dismisses both traditional discrete-point testing (as exemplified by Lado and Heaton) and integrated testing (as exemplified by Oller) as incapable of assessing communicative competence because they are not representations of 'natural' use of language: they ignore what ethnomethodologists call the 'etcetera principle', by which speakers omit many things for economy, and the 'ad hoc' principle, by which hearers make sense by filling in unstated propositions (26). She offers several suggestions for test types which might conform to her demands, but not all of them are 'natural' in the sense she has defined earlier, though all 'connect with the real world' in some sense.

Harrison (1983a) suggests three desiderata for a communicative test: it should assess language used for a purpose beyond itself; it should depend on the bridging of an information gap; it should represent an encounter. These lead on to the further suggestion that 'the answer to the question : Is it communicative? depends on the circumstances and the ability of the test technique to fit them' (79). He then raises four issues which he considers crucial in the discussion of communicative testing:

- content, which demands that learning, assessment and application should be recognised as in some kind of interplay with one another;

- groups and the individual, with the implication that communication requires more oral testing in groups and yet at the same time more concern with the individual's readiness;
- judgements, relating to decisions on what has been understood by the candidate and also to the problems in assessing live exchanges, including the importance of context; and
- realism, to include the nature of the task, the plausibility of the student's role and the authenticity of the language involved.

Rea (1985) points out (as Moller did in 1981) that the idea of relating language testing to communication is not new. She refers back to half a dozen discussions on the distinction between language competence and language performance, from J B Carroll onwards, and to discussion of situational appropriateness of utterances and communicative realism. Rea's concern is mainly with the integration of testing within a programme of instruction, that is, progress testing. For this, she outlines a system with a diagram (30), but there is no detail on the testing itself.

Swain (1985) on the other hand, explains at some length what communicative testing means in the context of Ontario's province-wide assessment of the communicative performance of immersion students at secondary level. She suggests that there are four principles of communicative test design:

- start from somewhere
- concentrate on content
- bias for best
- work for washback.

Starting from a theoretical framework (eg Canale & Swain 1980) 'assumes that a "scientific" rather than "evaluation" model underlies test design and implementation' (38); concentrating on content is implemented by providing information in a booklet on a theme (in this instance, summer employment for students) which promotes motivation and offers students information which is new to them and from which they may learn something; bias for best means giving the student favourable conditions, eg being allowed to work at their own speed and use dictionaries and other reference help; and working for washback means involving the teachers in the development work, and also in administration and scoring for practical purposes, since the scale of the enterprise demands it, and also because 'we hope not only to change aspects of what is taught but also to suggest alternative teaching-learning strategies' (44).

Skehan's survey of 1988, 10 years after Davies', covers new ground, for the key issues have changed. His five sections deal with the structure of general language

proficiency; communicative testing and ESP; developments in achievement testing; test method effects; and advances in statistical techniques (Skehan 1988:211). However, he considers that the debate over the conditions which make a test communicative does not appear to have made much progress: current reference is still to reformulations of the performance conditions set out by Morrow (1977 and 1979) (p215). But problems such as how a specific performance may be generalised and what performance tests imply are reported to be the subject of wide-ranging discussion and varied experiment. A crucial question is still that of generalisation or transferability, which was seen to be a problem ten years earlier (Morrow 1977). The way forward is for tests 'to be justified (and their generalisability established) by a mixture of theoretical and empirical work' and this should include the further development of performance tests, a special case of direct tests, a well-known example of which is the FSI interview and its derivatives (216-7). Skehan cites further examples of tests designed to test candidates' ability to cope with realistic tasks, but concludes that the problem has not yet been solved.

> The main question here is whether such techniques are likely to give measures for individuals which are accurate reflections of how they would perform when confronted with real-life communication.
>
> Skehan 1988:218

As to proficiency and achievement, Skehan separates them so far as to allocate them to different sections in his survey. Proficiency is what the construct discussion is all about; achievement is concerned with the growth and development of proficiency over time. This area includes longitudinal studies, coursebook progress testing and second-language acquisition research. Skehan thus takes a much more complex view of achievement than the usual flat description of it as the aim of tests (or examinations) at the end of a course.

Restricting his scope and at the same time accepting the challenge of a concept which, as shown above, has multiple interpretations, Weir (1988) writes under the uncompromising title *Communicative language testing*. His approach is to consider what communication involves, and he concludes that a variety of tests is now required, rather than a single 'best test' (11). A test within what Weir calls 'this communicative paradigm' might, he suggests, be concerned with:

- emphasis on interaction between participants
- unpredictable form and content
- purpose, fulfilling some communicative function
- relevant texts and authentic tasks
- a profile of performances in various contexts
- emphasis on appropriateness of language used
- direct testing methods

- qualitative rather than quantitative assessment of productive abilities (32-33).

On the distinctions to be made between performance and competence, Weir suggests that there are bound to be confusions. His views echo Rea's (1985), that competence cannot be evaluated except through its realisation in performance (Weir 1988:10).

Weir does not mention the terms achievement or proficiency in the 1988 book, but he takes a classic approach (no different from Lado's) in his book for language teachers: achievement should be firmly rooted in previous classroom experiences and proficiency is concerned with a candidate's ability to perform in a specified target situation (1993: 5,6).

A more theoretical attack on the problem of communicative testing is undertaken by Bachman (1990), who calls his book (after apparently much deliberation - see p ix) *Fundamental considerations in language testing*. The difference between this and Weir's commitment to the communicative is significant. Weir (1988) is interested in what benefits theoretical discussion may offer for the design of tests to be used with students in the field: Bachman is more concerned with further theorising about what communicative competence may be shown to consist of, which leads on to abstract discussions centred on construct validity. Bachman's brief discussion of communicative testing (1990:320) appears as an aspect of 'authenticity' among the 'persistent problems' for the future. He considers the concept to be summed up in the four criteria offered by Swain (1985), as quoted above - start from somewhere, concentrate on content, bias for best and work for washback, with the addition of the notion of an information gap, as suggested by Morrow (1979) 'and others'. He supports the idea of a thematic integration which relates a series of tests to a common subject (eg Wesche 1987) or to a source booklet (eg ELTS and TEEP). His overall view however (with Shohamy & Reves 1985) is that the only purpose of the interaction between tester and testee is to obtain an assessment of the testee's performance (20), which, though true in a literal and myopic sense, is not far from the traditional cat-and-mouse game of examiner and candidate trying to outwit each other through the medium of the examination paper (Harrison 1979:7) and a long way short of Swain's principle of 'bias for best'. Bachman's contention that oral examination interviews have their own illocutionary authenticity is illustrated from his own experience, but sudden switches of topic and role, as he practised them on the occasion he quotes, do not enhance the value of exchanges in tests as communicative events, and deserve the candidate's expostulation: '"You crazy!"' (1990:321).

Bachman's main contribution to the language testing debate is his further development of earlier formulations of competency, especially that of Canale & Swain (1980) and Canale (1983). He sees communicative ability as a complex of competences, knowledges and mechanisms, with *strategic competence* in a central position and *context of situation* out on the periphery (1980:85). He claims that this

structure builds upon the empirical findings of a study by Bachman & Palmer which suggests that 'the components of what they called grammatical and pragmatic competence are closely associated with each other, while the components they described as sociolinguistic competence are distinct' (86). This seems a fairly flimsy foundation for the complex and wide-ranging structure which follows, and in the event is not directly reflected in it. The diagram illustrating Bachman's concepts (87) shows *organizational competence* and *pragmatic competence* as subsidiary to *language competence*, and at the same time leading to two more competences each (*grammatical competence* and *textual competence* under *organizational competence*, and *illocutionary competence* and *sociolinguistic competence* under *pragmatic competence*). Out of these four competences fall 14 other lower level concepts. He follows Canale & Swain (1980) in labelling all the mechanics of language together as 'grammatical', adds *cohesion* and *rhetorical organisation* under *textual*, and then sets out four functions under *illocutions* and three sensitivities and a cultural factor under *sociological*.

This is a sizeable labelling exercise, intended to set out a comprehensive description of what the overall concept 'language competence' may consist of, but it is not clear how all these dimensions can be taken into account in the assessment of an individual student on a specific occasion, though they might conceivably be applied in practice across multiple assessments and on various occasions. The attempt at total coverage suggests comparisons with Munby's specifications of needs (1979), and seems likely to have the same stultifying effect in implementation.

Bachman's view of achievement testing is given in one short paragraph.

> These tests, or quizzes, which are based on the content of the course, are referred to as achievement or attainment tests. ...While the specific types of tests used for making decisions regarding progress and grades may very greatly from program to program, it is obvious that the content of such tests should be based on the syllabus rather than on a theory of language proficiency. That is, they will all be achievement tests.
>
> Bachman 1990:61

It is clear that at this point Bachman's main interest lies in concepts rather than in practicalities. He claims that there are two types of definition of language proficiency (41-2): the first says what the learner can do, the other says what knowledge he possesses. The first is operational, defining how the learner applies to situations what language he has; the second is componential, defining a set of abstract abilities.

In spite of quoting Morrow (1979), Swain (1985) and Wesche (1987) among others who have already developed practical examples of communicative tests, Bachman seems to consider that communicative approaches to language testing have not yet borne fruit, for he calls for a combination of the second and third trends suggested by Spolsky into a 'psychometric-communicative' trend. This will 'appropriately apply the tools of psychometrics and statistics to both the investigation of factors that

affect performance on language tests and the development of reliable, valid, and useful tests of communicative language abilities' (299). But if there is the incompatibility which later writers (eg Gipps 1994) find between newer kinds of educational assessment and traditional forms of reliability estimates, this hope seems likely to remain unfulfilled.

Bachman's exploration of competencies may have been so thorough as to have exhausted interest in competency theories: perhaps there is little more to be said on the subject. But McNamara (1995), after discussing what performance may mean by reviewing a range of abstract models of competence from those implied by the FSI tests of the 1950's to those explicitly described by Canale & Swain (1980) and Bachman (1990), ends with the argument is that although progress has been made in the description of the components which go to make up language performance, there is more work to be done on the modelling of non-linguistic factors, which is so far inadequate. He seems to be looking for new themes to explore, and offers a new angle on learner interaction by pointing out that not enough attention is given to what happens between participants in language performance, the present concentration being almost entirely on the activities of the individual candidate. In a later paper, he finds a wider area of interest in the ethics and social impact of assessment, which has been affected by 'a profound theoretical reorientation' (McNamara 1998).

This theme is explored by Gipps (1994) from the starting point of a 'paradigm shift' in assessment, 'from psychometrics to a broader model of educational assessment, from a testing and examinations culture to an assessment culture' (Gipps 1994:1). The broader model has several strands relevant to the present discussion, the first of which is a potential connection with a communicative approach. It was pointed out in the early days that attention to exchanges of meaning implied the need to accept multiple, individual interpretations, so that the traditional emphasis on reliability was no longer adequate to cope with the results of assessment based on social appropriateness rather than ineluctable accuracy. Morrow (1979) made the point in relation to the development of the RSA test (as it was then); and Underhill (1987) attacked 'testing experts' for insisting on traditional measures of reliability for speaking tests. Both were criticised in their time for polemic against received wisdom (Alderson 1981, Fulcher 1990), but now current thinking seems to have caught up with the notion, though not the implementations, they suggested.

A second strand in the new discussions is the notion of evaluation as a means of obtaining information about existing educational processes, or even bringing about developments - Rea-Dickins & Lwaitama (eds) (1995) - with the assessment of students as an important but only contributory element (Rea-Dickins & Germaine 1992). This perspective recognises the importance of interested parties beyond learners and teachers, such as school administrations, parents, employers and any individuals or agencies who are part of the context or are to be considered as

'stakeholders' (Gipps 1994, Rea-Dickins 1997). This viewpoint leads on to a third strand, a political/social interest in accountability, which looks for value for money in educational commitments and results in pressure on examination results. This is countered by practitioners' demands for more open and fair assessments over time, with portfolios of work and statements of achievement rather than 'sudden death' examination papers (eg Brindley 1998). These effects are observable not just in Britain but also in the United States and Australia (McNamara 1998, Brindley 1998). The difficulties in allocating social expenditures are compounded by movements of people in search of work so that language assessment becomes an aspect of immigration policies (Hawthorne 1997), and by the demand for flexible training for short-term employment (if any). At this point, the ethics of testing comes into focus, and in language assessment this results in a special issue of *Language Testing* in which Spolsky (1997) gives a historical perspective to 'gatekeeping', Lynch (1997) searches for the ethical test and Davies (1997) sets out demands for professionalism in language testing.

Finally, there is renewed interest in what empirical evidence the learner can offer beyond contributions to introspection studies (Færch & Kasper (eds) 1987). For example, the content of scales, rather than resting on speculative descriptions written by testers, may be better derived from learner responses (Chalhoub-Deville 1995, Fulcher 1996); learners may arrive at responses differently for tests considered to be equivalent (Douglas 1994, Shohamy 1994); learners at interview are likely to be treated differently because of varying 'supportive practices' in interviewers' reactions to learner responses (Ross 1992, Lazaraton 1996). This concern with individual variation is another aspect of the fairness debate: how to ensure that assessments report accurately on the learner when there is increasing pressure on results.

Summary

These current discussions are not explicitly about 'communicative' assessment. They take communicative aims for granted and go beyond them to investigate what society is demanding of assessment and what implications this has for theory. The interest of the present work is more restricted: it considers only what impact theory may have on practice so as to inform a specific investigation of the relationship between learning and subsequent assessment. For this, some summary is required of the above account of theories, from Lado to McNamara. Several themes recur, and relating these to the eight principles appearing in the review of communicative methodology of Chapter 1 could be useful in showing how far the terms used in learning and the terms used in assessment may be congruent. A count of topic references in the discussion so far shows that, in approximate order of frequency, the recurrent testing themes are:

- *authenticity* (to include such concepts as realism and natural use)
- *context situation* (information gap, appropriate use, circumstances)
- *performance* (interaction, groups, bias for best, can do, profiles)
- *context language* (competencies, washback, awareness)
- *needs* (prior analysis, purposes)
- *integrative* (global)
- *strategies* (processing, non linguistic, prior knowledge)
- *qualitative/quantitative* (judgements, criterion reference, test method)
- *generalisation* (representative)

Of the writers on communicative testing who have been quoted above, the least concerned with the methodological principles is Oller, who has his own approach based on his independent theory. He is not interested in realism, only representativeness, and in this he is in company with Alderson (1983), who maintains that communicative tests as such are not needed, and Davies, who considers that communicative language testing does not require that the tests themselves should be communicative (1986:61). But Oller's promotion of integrative testing, even though resulting in indirect (and, to some commentators, inadequate) tests, is consonant with the general view that communication always involves more than one language skill, even though this principle is not necessarily put directly into practice in tests. For Skehan's review (1988) interest also lies elsewhere, in that he is reporting on development since the previous survey on testing, and his finding is that there has not been much advance in ideas about communicative testing over the previous ten years. In company with Alderson & Weir, he is concerned about *generalisation*, the extent to which a communicative test, given its identity with context (which is in principle specific), can be accepted as representing performance on other occasions. Those referring most frequently to the eight principles are, perhaps not surprisingly, Morrow (who had a major part in formulating them) and Harrison (who identified them in Morrow & Johnson's writings for the purposes of Chapter 1). The principles most often mentioned by both methodologists and testers are *authenticity, needs, context* (of both language and situation) and *strategies* (coping with situations and interactions, both 1:1 and in groups). The correspondence between these accounts of communicative concerns in learning theory and testing theory is otherwise somewhat indirect, implying that at an abstract level of discussion, the concepts associated with communicative teaching/learning and communicative testing/assessment are some distance apart. On the other hand, testers have their own concerns, including such concepts as *performance* and *quality/quantity*, though materials writers and teachers might well benefit from considering them with their own aims in mind. But the differences in wording, and therefore apparently of view, between methodologists and testing specialists may be merely the result of differing discourse for essentially similar core ideas, since there seems to be little intellectual contact between the two groups, except in the person of Morrow.

At the start of this chapter the difference between achievement and proficiency testing was foreseen as a problem for the present work. Various commentators' views have been reported on at intervals during the above discussion in an attempt to find some consensus of opinion, and the arguments condense into two dichotomies. The first is distinguishing *achievement* testing as the assessment of specific, taught skills at points within the course and at the end of it, from *proficiency* testing as the assessment of overall abilities as applied to the requirements of the outside world. Parallel with this distinction is the difference between competence and performance, which is often discussed but usually as reflecting a separation between the definition of a theoretical construct for academic purposes on the one hand, and on the other, all kinds of examination and test. Bachman confuses the issue by dividing the theory which is to form the basis of proficiency testing into operational and abstract abilities. Harrison confuses the issue further by maintaining that the assessment of proficiency may well centre on the use of language for practical purposes, but the language which is so used has been methodically learnt, mainly in a classroom, so that elements of achievement must also be included. A solution to these nomenclature/concept convolutions is offered by Canale (1985), who suggests the concept 'proficiency-oriented achievement testing', which he defines as 'the systematic study, within a specific course or program, of students' progress in actually putting to use what they are learning, i.e. their progress in transferring controlled training into real performance.' After such lengthy rumination, it is with some relief that this definition can now be accepted as adequate for the purposes of the present work, with the understanding that henceforth the word 'achievement' carries these proficiency-oriented implications.

2.2 Applications

The review of theories, opinions and attitudes which has occupied this chapter up to this point does not carry the argument forward far enough for practical decisions to be made about what is to happen in actual tests, beyond ensuring that as far as possible any materials should fulfil the criteria described above as 'testing themes'. However, as theorists constantly maintain, there needs to be a theory on which practice is to be based. The question now is therefore how this theory is to be achieved, if it is not to be by means of propositions derived from numerical (probabilistic) data. Propositions of this kind are to be ruled out because any data collected for the purpose of deciding on the communicativeness (or otherwise) of tests are achieved by setting students tests which are by definition not communicative, since they cannot have been set according to an as yet unstated theory. If communicative tests are different in kind from 'traditional' tests, neither construct nor concurrent validation may be based on comparison, even if it builds up through the complex correlations of factor analysis. An alternative is to reverse the approach by studying existing tests which have some claim to communicativeness, with the intention of discovering what their particular

implementations of the above criteria may imply for the present work. The results of this study are likely to suggest problem areas which need further discussion.

The selection of tests or test systems as representing communicative testing in practice is problematical because there are far too many to be considered. For example, Jones found that in 1982 there were already over 100 speaking tests in existence (1985:77). One solution to the sampling problem is to find examples which illustrate different kinds of approach to the assessment of communicative ability and then apply to each example the list of criteria derived from the reporting of §2.1 above.

Sixteen tests (or testing systems) were investigated in the light of these criteria. There are ten which are general in application and widely available, though aiming at a known candidature, two developed as tests for specific purposes and four of an experimental or investigative kind. Inevitably there are constraints in every case which limit the realisation of communicative principles, however far it was envisaged that they would be implemented. The investigations are intended to show what has been achieved in the development of testing procedures relevant to commonly held views on the nature of communicative assessment.

An example of the information obtained for each test or testing system is given in *Figure 2.1*, which represents the syllabus of the English Speaking Board. Similar descriptions of the other fifteen are given in Appendix 2.1. Each description gives references, origins and characteristics of the test, and then summarises its response to the nine areas discussed above, ending with a summary of the main attributes of the test in relation to communicative criteria.

The results of this survey are reported in *Figure* 2.2, and a commentary then follows.

Figure 2.1: An examination syllabus with communicative elements

ESB

references: English Speaking Board (1994/5) *Syllabuses for EAL (English as an Acquired Language)*; Burniston (1968, 1982) *Creative oral assessment* (Southport: ESB)

origins & characteristics

1953. Established 'to promote & encourage all aspects of oral communication', as an improvement on the then current concern with elocution and correct speech production in examinations for public speaking and drama. Mainly for L1 speakers, but also 'a finely graded series of assessments for students for whom English is not their mother tongue...from the simplest introductory level Foundation through Intermediate to Advanced'. Oral test only.

authenticity 'to encourage enquiry, experiment, discovery & enthusiasm [which] involves the "reporter" in a variety of preceding practical written work & encourages dialogue with skilled adults in & beyond school & college', personal project as topic for presentation & discussion
context situation 'to give confidence in speaking English in a group situation...a supportive atmosphere is promoted'; 'to provide a meeting point of craft, trade, commerce, science, arts & leisure activities through oral, manual & visual presentation to a participating group'
performance task is to express own enthusiasm, explain, respond to questions on topic/artefact; individual written report from examiner to candidate
context language '[at earlier levels] considerable tolerance towards errors in grammar & pronunciation...the main priority is communication'
needs [no mention of analysis]
integrative 'not only effective oral communication but to heighten aural sensibility so that listening becomes an active disciplined experience'; 'all candidates will be expected to be active members of the listening group, asking questions & joining in discussion'
strategies 'candidates sitting in semicircle with the assessor as part of the group'; 'facilities for displaying candidates' pictures, drawings...must be provided'; 'at this level [Advanced] it is important that candidates are able to use strategies to overcome limitations of grammar & vocabulary'
qualitative/quantitative 'useful for examiner to assume that the candidate is "average" until he proves otherwise'; statements illustrate requirements for A-E on fan-shaped diagram; 'although we do not assess numerically it is useful to think in terms of' percentages for each section of the examination, with project weighted double; total aggregated in terms of 5 categories (pass, good pass, very good pass, credit, distinction) to arrive at final grade [guidelines on how to do this]; not summation of marks to a total; individual report
generalisation presentation and discussion skills regarded as transferable to everyday contexts; moderating: 'a wise examiner goes through marks with the teacher at the end of the day & discusses any anomalies'

attributes

- speaking as a key to personal development
- group assessment
- individual reports
- aggregation of judgements

Figure 2.2: Communicative criteria and tests

CRITERION

TEST	authenticity	context situation	performance	context language	needs	integrative	strategies	qualitative/ quantitative	generalise
ESB 1953	commitment to own project		present, explain, enthuse, answer questions; group			preparation in advance; speaking and listening discussion	as member of group	aggregation of judgements; individual report	
FSI 1956			managed conversation					attitude scale approach	
ARELS 1967		utterances for social situations						marked 0/1/2 appropriate rather than accurate	
OMLAC 1978		storyline	demonstrate what learnt		expanding social needs (speculative)				
T level 1979			reversal & repair in speaking					guide for speaking examiner from cand answers	
IELTS 1980		source booklet (abandoned)			profiles (speculative) (abandoned)		specifically assessed in speaking test	descriptive band scores	on demand; pretesting in operational
CCSE 1981	real texts	relevant tasks	performance criteria published (profile)				degrees of skill	certificates for 4 skills separately	
APU 1982/3			interrelation function/task; scales			assess listening through speaking	orientation to listener	appropriate rather than correct	

CRITERION

TEST	authenticity	context situation	performance	context language	needs	intergrative	strategies	qualitative/ quantitative	generalise
TEEP 1984		source booklet	profiles		researched			descriptive grades	
Avlp 1985	students supply theme	source booklet		application of constructs		tasks on source booklet		mark schemes derived from answers	
CAE (London) 1986		storyline/ situation	task based; student commitment			tasks on scenario (GTE)		positive marking; statements	
Israel Matric 1986			range of tasks; group discussion				interaction a criterion		research into policy
Trinity College 1993	student chooses topic						a criterion at all levels	12 grades in 4 stages specified	available on demand
DEIC 1994	real texts	relevant tasks	targets set for each task			integrated skills for tasks		criterion referenced	
FCE 1996				intentions specified	user-led content				
HKCEE 1996			group oral			tasks needing range of skills	use of Walkman for listening		120,000 candidates

2.3 Commentary on survey of tests

The issue now is how theoretical concepts have been realised in practice.

The summary (*Figure* 2.2) represents the salient points of the tests and testing systems included in Appendix 2.1. Entries in this summary include only those aspects specifically mentioned in the literature which seem to add impetus to the progress of implementing a communicative approach to testing. As a general survey of innovation, *Figure* 2.2 (read vertically) offers spot checks on how much advance has been made over the last 40 years in the areas regarded as important by communicative theorists. The extent to which these examination systems appear to have introduced communicative ideas is indicated by a horizontal reading.

The patterns in *Figure* 2.2 are scattered. The blank cells indicate not that there is nothing to record, but that the test system in question does not appear to offer a noticeably communicative development in the relevant area. It would have been possible to find a comment for each cell in the diagram, but this would not have brought out the individuality of testing systems. For example, 'context language' contains only two entries because the use of texts (rather than sentences or short unconnected snippets) to supply context for the language being assessed is fairly universal, and so did not merit a special note as a realisation of theory. Of the two entries under this heading, 'application of constructs' (AvIp) is a uniquely stated aim, even though it may well underlie others' productions, and 'intentions specified', though common to many other tests, is presented as a new departure for FCE. Again, ARELS, though pioneering in its time, has only two special mentions, 'utterances for social situations' and 'marked 0/1/2, appropriate rather than accurate'. These instances underline the fact that the chart represents the researcher's choice of entries, based on a review of the literature which is inevitably selective.

The pattern of *Figure* 2.2 is nevertheless a useful starting point for some promising observations about communicative testing in general. The vertical readings should perhaps be discussed first, as being indicative of how far the theorists' criteria have been put into practice. The highest number of entries (12 out of a possible 16) appear in the *quantitative/qualitative* area. At the centre of this concept is the difference between norm- and criterion-referencing, though the distinction may be considered more of a cline than a dichotomy, or even two aspects of the same concept (Davies 1978), in spite of claims that in a particular situation the principles of criterion-referencing have been exactly applied. This is very rarely the case in the testing systems under review, if it means strict adherence to the rules set up in expert definitions. These expect a description of the objectives or domains which are to be attained, the setting of cut-off scores indicating 'mastery', and possibly a system of reporting back in the form of attainments rather than scores (Brown 1980:96). The criterion-reference/norm-reference problem is acute where language is to be tested in accordance with a communicative philosophy, since the 'usage/use' contrast (Widdowson 1978) is parallel with a difference between knowing and

applying, the former being appropriately assessed by cumulative scoring of individual units of knowledge, and the latter needing some more descriptive attack on the problem of how the results of an assessment are to be interpreted. An approach through the accumulation and manipulation of numerical data will result in probabilities; the description of what is to be attained and a statement recording whether it has been attained or not will result in representations. The first approach results in a definite score (or grade) which is likely to be less accurate than it appears to be; the second results in a statement which may be understood in different ways. It seems only fitting that the use of language itself, rarely less than ambiguous, should be assessed by means of an interpretative system of reporting.

The concept of *performance* (column 4 in *Figure* 2.2) relates to what the candidate is asked to do, or more accurately what he has been asked to show he is capable of doing, which is a step further from actuality and overlaps with the concerns of generalisation. But this keyword also includes the nature of candidate response, and since this is capable of almost infinite variation, it seems important to look for ways of operating mark schemes which can accommodate a range of acceptable responses, by setting up assessment criteria which go beyond the one acceptable answer which has been forecast and to which all candidates must conform to gain credit. This is a justification for a task-based approach which asks the individual student to deploy his language resources to achieve an end, often in collaboration with other students in a group, and this brings a new perspective to the idea of 'appropriacy'. This keyword also concerns whether or not performances are to be reported in separate categories, whether representing skills or competencies; and how far statements of achievement are feasible.

Some consideration of *context situation* (column 2) is essential in communicative testing, and is mainly achieved in the testing systems under review by the use of a source booklet, which includes a range of material which may or may not be used as a basis for the tasks to be attempted. But the use of a source book has implications for candidatures. In an ESP context, which is by its nature limiting, a source booklet is both helpful to test setters, since it shortens the search for materials, and also reassuring to candidates, because it places them in what should be a familiar world. But with a more diverse clientèle the relevance of the contents of the booklet comes into question. An alternative is to build the test sequence on a storyline which gives an extending context to the language being assessed: the longer this context goes on the more students learn about the situation and the more they can appreciate what 'appropriate' means. A further advantage of this strategy is that the more familiar the situation becomes, the less time is required for a candidate to settle into the necessary perspective.

In view of the importance attached to *strategies*, both as the evidence of a competence and as lubrication for the smooth running of a communicative exchange, they seem to be specifically mentioned more rarely than might be

expected, though a syllabus will often expect that strategies will be taken into account along with more central requirements. Other aspects under this heading include supplementary factors which are not directly within the language exchange, eg relationships with interlocutors (as member of group, as attentive listener) and the use of L1 (which is rare).

It is surprising that there are so few entries under the three headings *context language*, (which implies a more or less universal use of extended text as source, as explained above), *needs* and *generalisation*. There appears to have been a move away from the specification of needs, as exemplified in theory by Richterich & Chancerel (1977) and Munby (1978), and in practice by Carroll (1981), the only example of thoroughly researched needs being for TEEP, which was intended for a particular ESP audience. And this may be the reason for the decline in popularity of the concept of needs: tests, like course books, need to be aimed at as wide a clientèle as possible to reach viability - not necessarily just in financial terms, but also in reputation and credibility, so that aiming a test at too specific a candidature is counter-productive. But FCE, which might be expected to embrace communicative principles only with reluctance in view of its traditional status, is constrained by the conservatism of its clientèle rather than its own. UCLES conducted a survey among users of FCE to advise it on what the content of the revised examination should be, and the result is a different attitude, but many of the same actual formats as before. The content of the *generalisation* column is linked with sampling, and this is an area rarely mentioned in syllabuses and introductions to testing systems, though it must influence decisions about what is to be included and what omitted. Even if not referred to, it may loom large in initial consideration of target groups.

Horizontally, the most frequent entries overall are for ESB, IELTS, CCSE, Avlp and DEIC (5 each), and the least, FSI, ARELS, T level and FCE (2 each). In the case of ESB, this is not surprising because the Board was early in recognising the importance of personal commitment from the student, and hence commitment of the examination system to the individual. The ESB system is centred on discussions starting from a student's interests, prepared extensively in advance, presented to other students (and to the assessor as a member of the group) with artefacts and visual illustrations as available. An important by-product (which would now be called backwash) is the influence of the research and presentation practice demanded by the system on the individual's confidence and self worth. Enthusiastic and convincing accounts of her experiences as participant - examiner, group member, observer - are given by Burniston (1968, 1982), and sentimental as her accounts may sometimes seem, the fact is that her commitment to the individual does lead in logical sequence to the implementation of many tenets of the communicative approach before it was ever discussed as a new development.

The introduction of new tests is constrained by many factors, but a serious retarding influence on possible change is the administrative demands of the testing

organisation, particularly in the matter of candidate numbers and location. The case of IELTS is typical. Several aspects of the original ELTS (1980) which attempted to put communicative principles into practice, for example source books and the effort to set up needs analyses (however speculative in the event - see for example Clapham 1981) have been abandoned in IELTS (British Council et al 1989, 1997). The extension of the range of intended candidates from post-graduates and undergraduates to include not only technical trainees (which had been envisaged from the beginning), but also secondary school students and immigrants, has evidently demanded too large a library of interest-oriented source books. On the other hand, the use of bands for the reporting of results has remained, largely because they gained understanding with the establishment of ELTS and have become an accepted international currency among users.

The Royal Society of Arts' CUEFL was developed from Morrow's (1977) theoretical paper on communicative testing, and it is therefore to be expected that CCSE, its successor, would be mentioned fairly frequently in *Figure* 2.2. Morrow's desiderata, which were discussed earlier in this chapter, are in the main fulfilled, but his emphasis on the problems of generalisation, as chiefly a matter of sampling, does not seem to have moved further forward than the specification of a syllabus for a wide general audience of 'adult users' (another example of a constraint imposed by the examining board so as to encourage high numbers for viability?). But at least there are detailed statements of the degrees of skill required for each 'area' (Reading, Writing, Listening, Speaking) at each of the four levels. The most striking departure is from Morrow's original requirement that 'performance tests are necessarily integrated'. CCSE (& CUEFL before it) offer certification in the four skills ('areas') independently, so that candidates may take some or all of them at one time and some at different levels from others. This is an admirably flexible arrangement, but undermines one of the main arguments in communicative polemic, that the exchange of meanings between two people (with very few exceptions) involves more than one skill. It may be that when Morrow talks of 'integrative' he is contrasting it with 'discrete point', meaning only that texts are acceptable but isolated bits of language are not, but if so this is something of a prevarication in the face of a change of theoretical stance.

Another test system allocated five comments in the chart is Avlp. Here, a source book is appropriate, for the candidates are a homogeneous group: secondary school students on 'immersion' courses (ie learning school subjects through L2). And it provides for the implementation of several communicative ideas, for not only is it derived from student preference in topics (with their consequent interest and impetus towards the completion of tasks), but it gives the students an impression of authenticity, and allows integrative use of skills. Finally, this is the first explicit reference in these test systems to allowing candidates' answers to shape the marking, which means that instead of setting targets which candidate performance is to reach, the criteria set for assessment take account of the communicative value

of student responses at pretesting. The same principle has been persuasively argued and successfully applied in secondary examinations in History (Southern Regional Examinations Board 1980, 1981). (This is categorically different from the well-established procedure for pretesting objective test items, which produces numerical data on which the value of items for future use is to be judged.)

The Institute of Linguists has always been concerned with language performance in practice rather than display of knowledge about language, so that its step into a communicative approach has been a relatively short one. Its latest syllabus in English as a foreign language (DEIC) is not surprisingly among those with five entries in *Figure* 2.2. It is the only scheme which provides a dossier of photocopied real material for candidates to work with in the completion of relevant tasks, with a deliberate mix of skills, again in a realistic situation. Explicit targets are set for each task in the form of a brief for the candidate, and examiners work from an assessment checklist which must be fulfilled for the candidate to pass. There are three modules and a background knowledge paper, all of which may be taken separately, but the Diploma is awarded only if they are all passed within a specified period. Additional requirements are laid down for the award of Distinction. Of all those recorded in *Figure* 2.2, this scheme approaches most closely the conditions for criterion-referenced assessment, as described for example by Brown (1980).

2.4 Consequences

The uneven relationship between theory and practice which has appeared in *Figure* 2.2 is now to form the basis for further discussion on central problems in the development of assessment systems which not only follow communicative principles but also relate directly to preceding classroom events. Three areas have been identified as problematical, and the background to each of them and its impact on the present work provide the substance of the next chapter.

Chapter 3

Problem areas

3.0 Introduction

The responsibility of this chapter is to investigate topics which are likely to cause problems in the assessment of language performance. The labels attached to the areas discussed below - tasks, levels and judgements - are to some extent arbitrary, but they allow a range of issues to be conveniently gathered under headings. 'Tasks' is concerned with what the student is asked to do in classroom activities and assessments, the central problems being the nature of a task, its relationship with 'real-life' demands, and what it expects of a student. 'Levels' is concerned with decisions about learning syllabuses and test specifications, the central problems being sequencing and the justifications for it. 'Judgements' is concerned with what decisions may be made about the value of a student's engagement with class work and assessment material, the main problems being the interpretation of relevance in responses and the nature of the information which is to be carried back to students and teachers.

These three areas link back in various ways to the criteria illustrated in Chapter 2. For example *authenticity* has considerable bearing on tasks in its concern with real input, realistic interchanges and relevant output; it also relates to levels, for example in the simplification or otherwise of input texts; and it concerns judgements in so far as marking systems need to consider the acceptability of responses for real world use. In addition, running through all three areas, is the notion of generalisation: it is important that any assessment systems proposed should widely applicable in principle, independently of specific learner interest or particular school circumstances.

The following three sections visit each of the topics in turn, referring to the nine communicative criteria as they appear relevant, and the chapter ends with a summary of the conditions which will need to be addressed in the development of the present project.

3.1 Tasks

The difficulties to be encountered with the concept of 'task' begin with problems of definition: what exactly is the nature of a task as used for assessment purposes? What differentiates one task from another, especially when they are linked in a sequence? How 'real' can a task be when it is done in a classroom setting? How

will learners know what a task expects them to do?

Definitions

Considering classroom activities as 'tasks' is commonplace, since any work done by any student can legitimately be so labelled, from looking up words in a dictionary to completing extensive assignments as homework. The importance of the concept for present purposes is that communicative language learning and assessment expect students to learn 'how to do things with words'. With this memorable phrase of Austin's (1976), a philosophical argument about performative verbs is extended by Wilkins (1972) and others into a discussion of practical applications for functional uses. The fundamental idea of a communicative task is that it goes beyond the limitations of the standard 'PPP' (presentation, practice, production) lesson which is reviewed in detail by Skehan (1998a:93-5) as an all-pervasive classroom procedure, and described by Rockwell (1998) as an inadequate teacher-training convention.

The demand of a task which is additional to the PPP procedure is that it requires a personal interaction between learner and text in the broadest sense. 'Text' here includes both spoken and written language and the production of other learners as well as teaching materials of whatever kind, including visual material. There is also a requirement that the learner should be challenged with a problem which is not a mere manipulation of language, and this in turn means that he will need to concentrate on meaning (the what) rather than form (the how), and reach some kind of solution in collaboration with (or in spite of the opposition of) other producers of text. A task such as this gives the learner the opportunity to apply what he has learnt to a realistic purpose, and at the same time involves him in an encounter with other minds which widens his knowledge of the language and how it can be used. These conditions introduce the notion of the individual learner both contributing to and benefiting from involvement in a task. In brief, a task involves personal commitment and collaboration with other learners and at the same time extends the learner's experience.

But defining exactly what is meant by 'task', in this sense of an opportunity to apply learned skills, is problematical. A reference to Piaget (1926) may help. The essentials of a task in Piaget's experiments were: a starting point which was of interest to the children and at their level of understanding; a problem which contained specific elements to be achieved; and an open approach to the content of the response, provided the specified elements were included. These three conditions (mutatis mutandis) are equally important in the setting of communicative tasks for language learners, as will become apparent in later discussion.

More recent and more topically pointed definitions of 'task' range from the convoluted to the practical. Candlin's 'working definition of a language-learning task' reads:

> One of a set of differentiated, sequenceable, problem-posing activities involving learners and teachers in some joint selection from a range of varied cognitive and communicative procedures applied to existing and new knowledge in the collective exploration and pursuance of foreseen or emergent goals within a social milieu.
>
> Candlin 1987:10

This complex attempt at a comprehensive statement which will define a universal application of the concept results in a formula which is so wide in its application that it no longer has much specific gravity: it becomes merely something that may or may not happen in the language classroom, or beyond it.

Nunan describes a more focussed equivalent, 'the communicative task', as

> A piece of classroom work which involves learners in comprehending, manipulating, producing or interacting in the target language while their attention is principally focused on meaning rather than form.
>
> Nunan 1989:10

This definition is general as far as activities is concerned, but has one important element: the attention of the students on meaning, 'getting something done', with language as a means to an end rather than as an end in itself.

Another, wider ranging, definition offers details on how a task is set up in the language class and what it requires of learners:

> "Tasks", in the context of language teaching, are genuine or simulated situations containing the features of real normal communication, in which learners have to apply their ability to use language to achieve the objective inherent in the situation. ...Accomplishing the task should involve learners in using language to overcome information gaps, react to what they have said or written to them appropriately, and to display to the full their attainment of flexibility and variety.
>
> Heyworth 1984:6

In spite of its odd wording, this definition is helpful for this discussion because it raises further questions, such as the difference between 'genuine' and 'simulated', the nature of an 'objective inherent in the situation', the necessity for an 'information gap', and the ways in which 'flexibility and variety' may be demanded by a task.

Skehan summarises various aspects of task-based instruction (1996, 1998), and then proposes a definition which sums up the views of a range of writers:

> a task is an activity in which:
>
> - meaning is primary;
> - there is some communication problem to solve;
> - there is some sort of relationship to comparable real-world events;
> - task completion has some priority;
> - the assessment of the task is in terms of outcome.
>
> Skehan 1998:95

All these examples of definitions seem to show that encapsulating the idea of 'task' in a simple formulation is a surprisingly complex undertaking. The difficulty lies mainly in finding a general enough definition to fit all circumstances. A request from

a teacher to a learner about the opening or shutting of classroom windows fulfils the requirements of Skehan's definition, but seems too trivial to count as a language-learning exercise. And yet tasks of this kind are included in Trinity College's examinations at the lowest level (Trinity College 1994), in a series of examiner demands which nevertheless fails to constitute a 'communicative' test because the requests being made are unconnected with any overall context and appear to exist solely for demonstration purposes. At this point, some further investigation seems necessary to help decide what activities may qualify as tasks.

Task types

One attempt to differentiate tasks according to their relationship with the real world suggests two general categories of situational task: simulations and puzzles (Harrison 1982b:405). A simulation in this case is based on the premise: Suppose you were... (in some real situation which could require your language skills). A puzzle is based on the premise: Here is a task... (with a problem to be solved, not necessarily one related to your eventual use of language). But the concept of a simulation has had wide application, and equally wide definition, from the field exercises of Army officer training to short interactions in classrooms. In the EFL field, the benefits of a simulation are generally considered to be much the same as those of role plays, and the two activities are often linked together as variations on one kind of activity. This occurs both in theoretical discussions (eg Wilkins 1976a, Dickinson 1981, Scarbrough 1981, Crookall 1984) and in practical suggestions for teachers (eg Herbert & Sturtridge 1979, K Jones 1980, 1982 & 1985, Golebiowska 1990, Porter Ladousse 1983 &1987, di Pietro 1987). Sets of material for role plays and simulations are published for classroom use, sometimes with input materials in photocopiable form (eg Lynch 1977, Hicks et al 1979, Brims 1982, L Jones 1983 among other more recent examples, many of them now incorporated in course books).

Various kinds of classroom interaction need to be differentiated in order to suggest which may be considered 'tasks' in the sense being investigated here. Role plays involve participants in some event, but as other people, not as themselves (Porter-Ladousse 1987); drama is a larger-than-life presentation of fictional events (Maley & Duff 1978, Butterfield 1989); case studies come from real life but do not typically involve decision-making; and the in-tray exercise of management training seminars is usually a lone activity without group discussion. Some of these conform to Skehan's definition, but none fulfils the conditions for individual involvement set out above: personal commitment, collaboration and extension of experience. A simulation however has particular and promising attributes. The essence is action: whatever documents may be used for input, whatever roles are allocated, the most important point is that participants work together to come to conclusions about the issues presented, making decisions, responding to others involved, and reaching some sort of end - both a purpose and a closure in the Gestalt sense (Clarkson 1989). It is important for learners to know that they have achieved something by

taking part in the simulation and at the same time feel that they have arrived at a satisfying conclusion. One further essential for a simulation, which Jones (1982) includes under 'motivation', is that participants will be making real decisions on the basis of the responsibilities they have undertaken, the information they have been given, and their own knowledge, attitudes and character, which they inevitably bring to the situation, so that they are committed to personal involvement in the action. This suggests that science fiction and other fantasies are not strictly within the rules, although Jones includes them provided the result is personal commitment to discussion and decision, even if not in a realistic context. Jones sums up his account with the short overall statement: 'a simulation is reality of function in a simulated and structured environment' (Jones 1982:5). Further discussion of the theory and practice of simulations is provided by Taylor & Walford 1978, K Jones 1980, 1985, Wright 1980 & 1981, L Jones 1983 and F R Jones 1991.

All the activities discussed so far can be considered to be variations of the concept 'simulation' because they have some relationship with the real world. The concept 'puzzle' encompasses any interactivity which does not claim to represent life outside the classroom. One of these is 'game', an idea which has an extensive range of connotations. Wittgenstein's concept of a 'language-game' is 'a more or less complicated shared human activity which might, or might not, have a utility which could be grasped and stated outside the game' (Kenny 1975:168). This concept of 'game' lies at the philosophical end of the range, with Berne's psychological interpretation of 'games people play' (1966) somewhere in the middle and Rixon's suggestions on how to use games in language teaching (1981) at the more literal end. This last is the usual meaning attached to 'game' in the present context, as a classroom exercise which stimulates useful activity with language but is not related to real-life decisions (eg Lee 1965, 1979, Byrne & Rixon 1979, Wright et al 1979). But even in this restricted sense, a game is still a 'shared human activity', which exemplifies the essentially dual (at a minimum) nature of language use, and at the same time is an opportunity for individuals to contribute aspects of their individuality to a common cause. Other kinds of 'puzzle' are information gap exercises (eg Geddes & Sturtridge 1979, 1982, Klippel 1984, Ur 1981); activities involving the commitment and revelation of self (eg Moskowitz 1978, Brandes & Phillips 1979, Porter Ladousse 1983) and tricks and manipulative problems labelled literally 'activities and puzzles for language learners' (Maley & Grellet 1981).

Tasks in language learning and assessment have both linguistic and phenomenological aspects. In terms of the communicative approach, the idea of 'task' is consonant with doing something constructive with the language. Any activity in a language lesson might be termed a task, but the communicative assumption is that mechanistic learning, for example sense-less repetitive drilling, can contribute little (if anything) to the actual learning process (Dakin 1973). The point of an interactive task is that it represents a situation involving others, so that there is an external purpose in mind - the application of language to situations presenting problems.

These problems can be considered to be of two kinds: those inherent in the situation and those arising from the linguistic demands made of the learners involved, who grapple with them with mind and language. The linguistic problem is succinctly described by Wilkins:

> The ability to perform is a matter of selecting accurately from the repertoire of language possessed by the individual at that stage (lexicon, syntactic rules, phonological rules, etc.), and of carrying out the complex linguistic tasks and sub-tasks involved in acute, temporal synchrony.
>
> Wilkins 1983:35

At the same time, task-based learning and assessment, when related to situations external to the learning process, are consonant with the daily acts and aspirations of human beings, coping with the world as they experience it in the present. Giving learners something relevant to achieve provides a motivation beyond the use of language. Some principles of Gestalt psychology suggest useful guidelines:

> We tend in perception to complete what is incomplete, just as in behavior we tend toward a situation which is completed and feel tension until it has reached some conclusive stage. This is the law of closure...
>
> Rivers 1964:179

In more philosophical terms, a relevant task fulfils the Gestalt principle of offering the possibility of choice.

> For existentialists the notion of 'existential choice' is fundamental to being human. This means that each of us is choosing what we accept, reject, think, feel or how we behave.
>
> Clarkson 1989:13

Reaching the end of a task will engender a sense of achievement; what Gestaltists term the satisfaction of a need. What is required for present purposes therefore is some means of offering learners the opportunity to use the *language they have available* to solve a problem which is *worthwhile from their point of view* so that it satisfies their need for active engagement in both fields.

Authenticity

True authenticity, however it may be defined, is impossible in the classroom, other than in teacher/learner organisational exchanges. Davies calls it a 'chimera' (1978:225). But the problems inherent in the relationship between a task undertaken in a classroom and a real-life event must be solved if the principles of communicative learning are to be carried through into classroom testing. What the test theorists say about it has been discussed in Chapter 2, and its practice has also been touched on, with the intention of exploring the nature of authenticity in testing: here, the concern is what makes a task authentic and how authenticity, as far as it exists, can be promoted in devising tasks, with special reference to assessment.

A broad perspective suggests that the best that can be done is to ensure that attention is given to the origins of input and the context of the students' future use

of the L2. In spite of the varied authenticities suggested by Clark (1987), the essentials for an assessment task are only three: how nearly the stimulus material approximates to a real-life manifestation; how nearly the action demanded by the task reflects the requirements of a real context; and how far the task enables (or requires) students to find language appropriate to that context.

A variety of authenticities is suggested by Breen: 'In language teaching,... authenticity of data, authentic communication, authenticity to the state of the learner, and authenticity to the classroom itself have to exist in a state of interplay.' (1983:59). Widdowson goes further, writing of 'authenticity not as a quality residing in instances of language but as a quality which is bestowed upon them, created by the response of the receiver' (1979:165). This sets the problem in a different perspective: real texts, real tasks, real language are subservient to real communication. Evidently, for assessment purposes, authenticity will depend on judgements made about what happens in an interactive task, and this can realistically be made only about the actual, identifiable responses produced by each student. But the multiple nature of these responses is such that they need to be approached from several angles, to include not just fluency and/or correctness, use of strategies, appropriateness of language to situation and the other desiderata proposed by the theorists, but encompassing also appropriateness in the wider context of participation: who says what in response to what (input from both task and other participants), how far involvement in the task brings each participant's creativity (imagination, inventiveness, idiosyncrasies) into action. Finally, it is important to recall that authenticity in all cases also includes the context the students bring with them (which has been aptly termed 'baggage') and the contexts they find themselves in as learners, as well as to the context of the language they need to negotiate with.

Expectation and rubric

When an idea for a task has been prepared for learners to use, the question arises of how it is to be clearly explained to them, for there will always be intentions and implications which are not self-evident. This may involve an example or a trial run. In any case, the teacher can extend the explanation as necessary, using it as part of a learning process. He may even decide to give no preparatory information at all, as a deliberate problem within the task.

But rubric is a perennial difficulty for testing, since the candidate must understand what he is to do before the task starts. It should be possible to explain it to the learners in such a way that their response is not muddied by the mechanics of the task, but can be considered a clear indication of performance. At the same time, there are many other factors beyond mere instructions which impinge on students' attitudes and commitment, for example the importance for them of the outcome, the value they see in the tasks, even the nearness of the end of the course. Solutions include introductions in the students' L1 where this is possible, eg in homogeneous L1 classes; practice with formats and content, eg using books in

preparation for Cambridge examinations (which fill 12 pages in Longmans 1997 catalogue); worked examples at the beginning of each test paper, though these must strike a balance between comprehensiveness and time spent in absorbing them; grouping of questions into types, eg all listening before all reading, all objective items before more open-ended ones; and a thread leading through a set of tests so that the situation is already clear before a new section starts and does not need to be worked into afresh each time, eg a storyline, a source book. An exhaustive list of all the factors involved in rubric is attempted by Bachman & Palmer (1996:181-191), and this gives welcome if pernickety attention to an aspect of assessment which is sometimes taken for granted or at best understated in discussions about test writing. Heaton's more practical coverage (1988) on the other hand, though helpful in principle, does not deal with the particular conditions arising from the assessment of interactive language use. Long and colleagues, as quoted by Allwright (1988:153), refer to 'the frustration felt by practicing teachers at the apparent inability of their students to transfer knowledge of the target language to situations outside language classrooms'. And, it might be added, to situational assessment tasks within classrooms: students need to learn how to use what they know to fulfil the demands of a task, a transfer from fragmented learning to concatenated doing. This is an additional element in learning which was sometimes neglected in early 'communicative' materials but has been encouraged by the extended availability of resource books for teachers about applying the learnt to the situational, such as those referred to above. The difficulty of explaining adequately still remains.

Task and assessment

Throughout the above discussion, which has covered types of interaction which could be classified as tasks, the issue of authenticity and what is expected of learners, there has been no attempt to define the concept of 'task' for present purposes. The discussion has been mainly concerned with tasks for teaching and learning, and though the burden of this thesis is that learning and assessment systems should be more closely allied, there are important differences to be taken into account, beyond the mere addition of considered judgements to suitable classroom activities. Tasks for learning purposes may include many activities which are not necessarily communicative in intention. The aim of tasks concerned with improving listening and reading, or developing grammar and vocabulary is to build up the learner's repertoire of language without needing to be 'realistic' tasks in the sense expected by 'authenticity'. Tasks for assessment purposes on the other hand can restrict themselves to the application of what has been learnt, gathering together what has gone before, and applying it to a situation which will be as realistic as possible. Assessment can show students that they have been learning something they can use both now and later (inspiring motivation for future learning) as well as giving them information on what they have achieved so far.

For the purposes of the present project, an assessment task is to be defined as an

activity which

1 puts into practice preceding learning;

2 involves learners in productive interactions with text, including that produced by other learners;

3 takes into account *authenticity, contexts, performance, needs, integration, strategies, quantitative judgements* and *generalisation* (see Chapter 2);

4 gives learners helpful information about their standing, both in detail (diagnostic) and in general (achievement).

3.2 Levels

The difficulty with levels in language learning is that they cannot be linked into a logical order of progression. In terms of language, no word or construction is necessarily more 'difficult' than any other. Defining sequences in accordance with mental development (Piaget 1926) cannot apply directly to adult language learners; and the results of relevant research across the curriculum (eg Biggs & Collis 1982 in a range of subjects; Harlen et al 1977 and Shayer & Adey 1981 in science; CSMS 1980 in mathematics) do not offer enough parallels with language learning to be useful.

But pure linguistics is not helpful either. Chomsky is 'sceptical about the significance for the teaching of languages of such insights and understanding as have been attained in linguistics and psychology', including the results of studies in concept formation (quoted in Allen & van Buren 1971:152). Others, however, faced with the problem that a communicative view of language learning logically excludes the use of hierarchical assemblies of language elements, find it difficult to formulate answers to the problem of syllabus definition. Corder, maintaining in 1973 that no ordering is possible because 'the structure of language is a network of interrelated categories, no part of which is wholly independent or wholly dependent upon another' (Corder 1973:297), later modifies this view to allow for 'Teacher Talk' as a means of grading the language the learner is exposed to, and 'a programme in which the syllabus sets the learner increasingly complex and demanding communicative tasks to perform' (Corder 1980:87). Krashen argues that, from the point of view of Second Language Acquisition (SLA) theory, the main suggestions for the grading of input are that there is an 'average' order of acquisition of grammatical morphemes for English as a second language (1977); that input of the early learner should be made more comprehensible by the teacher's modifications (1982:64ff); and that the input should be just beyond the student's present capacity (1982:21). The first of these is taken up by Pienemann and others. Research into the order of elements in language acquisition has shown, Pienemann argues, that word order in German is a fixed sequence for some Romance language learners, and further, that morphosyntactic features appear in order for ESL learners with Vietnamese and Polish as L1 (Pienemann et al 1988). This is an interesting finding for theories of

SLA but unless it can be shown to hold good over a much wider range of language activity, it is not a particularly useful guide in the writing of syllabuses either for courses or for assessment.

A logical order of difficulty?

As traditionally applied in language learning, the idea of levels seems to develop in two directions: accumulation or extension. Accumulation depends on counting, on the addition of more material to a starting base, probably in a particular order, but certainly on some principle of increasing value, so that upper levels contain more in terms of difficulty or applicability, and in any case represent increased merit on the part of the student. Extension on the other hand relates to a widening focus of interest for the learner, influenced in principle by social needs and based on recycling of functions in ever more complex interactions. It is associated with the appreciation of context and so deals with the appropriateness of a student's response rather than its exactitude, allowing him to react as an individual to the demands of different situations.

Attempts to place the components of language learning in a cumulative order have a long history, from Comenius (1648, quoted by Kelly 1969) to Pienemann (1988) and beyond. Comenius suggests that dimensions might include the few before the many, the short before the long, the simple before the complex, the general before the particular, the nearer before the more remote, the regular before the irregular (Kelly 1969:224). One difficulty with these categories is that they do not allow for different kinds of learner. For example, adults may well appreciate an approach which narrows down from the general to the particular, but school pupils would be better served with a curriculum which began with the particular (as reflected in their experience - family, home, school) and worked out to the general. Another problem is that Comenius' order does not take account of the nature of language development, for example that the irregular is likely to be the most frequent and hence the most useful.

If frequency is taken as a criterion however, further difficulties arise, since no intrinsic ordering of vocabulary is possible, in spite of the efforts of Thorndike & Lorge (1944), who provided teachers with a word book of 30,000 words; West, who edited a General Service List (1953, though the work goes back to 1936); and more recently Hindmarsh (1980), who produced 'a graded word list for materials writers and course designers', with a 'particular orientation towards the First Certificate in English' (vi). As Mackey (1965:159-201) points out, the problems with word counts include what is regarded as a word, which texts are to be used for the counting and what exceptions are to be made to the literal results of the counting. For example, the French government sponsored the preparation of *Le français fondamental* as a guide for syllabuses, based on spoken French, derived from 163 conversations resulting in 312,000 words, of which those occurring 29 times or more were listed. In addition, the list includes useful words which do not necessarily occur often,

established by asking pupils in a wide selection of schools to list what they considered were the 20 most useful words under 16 topic headings (Ministère de l'Education Nationale 1972). But even this extensive trawl does not give equal value to the days of the week: Mackey comments that *dimanche*, *lundi* and *samedi* appear in the first thousand words, but *mardi* and *jeudi* in the second (Mackey 1965:182). How far one would have to go to come across *mercredi* and *vendredi* is not recorded. Wilkins, in his full account of the shortcomings of vocabulary lists (1972:112-118) instances 'Thursday' in the same way as a possible absentee from a frequency count. He adds that anomalies like this raise doubts about the assumption that underlies all frequency studies: 'If frequency does not reflect usefulness in this case, why should it be thought to do so elsewhere?' (Wilkins 1972:116).

The advent of computerised counting, which can produce corpora containing millions of 'running words' and in theory has no end, enables much more data to be consulted. Meara considers that it has made all previous listings out of date, especially since 'authoring programs exist which will allow the non-specialist to undertake his own analysis for his own particular purposes' (Meara 1980:223). But the problems are still the same, even with 20 million words, which was the count for the Birmingham corpus as early as 1987 (Carter 1987:12). What is a word? What kinds of text are represented? How does frequency relate to usefulness?

The arrangement of grammar in an order of increasing difficulty might seem more justifiable than the listing of vocabulary because hierarchies clearly exist in the system, for example differences in verb forms (from past simple to past perfect or from active to passive). But even here, there are problems. Four distinguished grammarians produced a general syllabus for teachers which is 'divided into six Stages ... in an endeavour to mark steps towards progressive levels of competence'. They somewhat apologetically admit that it is not statistically based, but 'It is on the basis of [the authors'] combined experience... that they have agreed where the limits of this book should be drawn and in what order individual items should appear' (Alexander et al 1975). 'They have taken into consideration such factors as apparent frequency of occurrence, productivity, general usefulness, progression from simple to compound and complex, and pedagogical expediency'(v). This is a unique and valuable reference, but still cannot solve the problem of what grammar is to be taught at what level to fulfil the demands of a given situation.

Practical guidelines on the use of both vocabulary and structure for the writers of graded readers are suggested by O'Neill (1987). They are divided into four stages, for each of which there are general guidelines, general criteria for vocabulary selection, a vocabulary list and a structure list. The first handbook (1968) was based on an analysis of coursebook material available at the time, but was revised (1987) to reflect some of the critical changes in language teaching and materials design which had taken place in the meantime, including the suggestion that students now get to grips with authentic English earlier than before. O'Neill is well aware of the

consequences of a communicative approach but is caught in a publisher's trap of having to specify what is acceptable in a graded reader at a particular level, as advertised. His main criteria are 'accessibility (within the general reading competence of the student) and motivational push (driving the student to read on and even to skip over some language which is too difficult)' (1987:1). But the impression is that overall, O'Neill is not entirely out of sympathy with strict control of input to language learning, based on 'general and intuitive judgements about what is appropriate and what is not. The guidelines... are the product of more than half a century of this kind of intuition'(1987:2).

This constant search for a logical set of rules for relative difficulty underpinned what Modern Language teacher-agitators of the 1970s and 80s called 'a defined syllabus', but the restricted coverage which resulted paradoxically led to texts, situations and marking schemes for the examination papers which were even more artificial than before (see Davidson 1973, Moys et al 1980, Page 1974, with counterblast from Tucker 1982). In EFL the urge towards comprehensive definitions has brought ever more plurally detailed objectives for course books (for example Swan & Walter's 'multi-syllabus' of 1985 which 'is based on a combination of eight different syllabus-inventories'). The defined or specific syllabus tends to be associated with exactitude and expectations that a prescribed body of learning will have been accumulated by the student for each level in turn. This may be justified as clarifying aims and achievements but does not chime with the plurality and open-endedness of communication as it occurs.

An attempt was made by van Ek to overcome this problem: his apologia for the T level syllabus insists that the intention is for it to be only one of many such syllabuses (as the T level test is in the same way only one of many possibilities - Groot & Harrison 1979), and that the content is therefore open to addition and amendment according to circumstances. There is nevertheless reference to frequency and the relative importance of individual words in the syllabus (van Ek 1973). His somewhat sophistical argument in defining the threshold level and at the same time justifying the lists is that it is 'the level on which the learner has the ability to communicate in those verbal communication situations in which it may be expected that, as far as vocabulary and syntax are concerned, the lexical and syntactic command provided by the items of the lists suffices' (van Ek 1973:102). In the same report, Wilkins takes a less prescriptive view. 'By considering first what the content of utterances is likely to be, it is possible to decide which forms of language will be most valuable'. He also proposes recycling at higher levels (Wilkins 1973:143).

A logical order of social needs?

Perhaps, instead of counting vertically, with the inherent difficulty of setting cut-off points which cannot exist in actuality, the syllabus writer could think laterally. The extension model of increasing coverage in language learning is dependent on social rather than pedagogic concerns, on the situations in which a language user may find

himself, on his needs in terms of language for the interactions likely to involve him, or his own interests, in the widest sense. The student's progress up any set of levels in this perspective is usefully to be considered as an expanding universe of language contacts, well represented by Trim's parallel with individual mobility. At first, the child crawls around the room, then a house, and then his social context extends in turn to the supermarket, mother's friends, his own friends, school, town, work, links with other networks in other cities and in other countries (Trim 1978:7). This image is also intended to illustrate the stresses of an adult L2 learner: frustration in situations which he is, like a child, incompetent to handle, but only because of inadequacy of language. This leads to the suggestion that instead of generalised objectives such as elementary, intermediate and advanced, 'one looks to providing a learner with the body of knowledge and the skills he needs to solve a defined communication problem' (Trim 1978:8). The necessary definition has been transferred from the components of language to the uses to which it will be put.

This is the logic behind the analysis of needs. If the reason for learning a language is practical rather than educational (though education is never far behind), both learner and teacher have to know what the practical ends are. Richterich & Chancerel propose an approach centred on the learner and his freedom of choice (1977:7). For the individual learner, they envisage a two-stage analysis, before the course and during the course, and in addition set out a series of different kinds of information which should be obtained to implement the principle of needs, for example by the learner, the teaching establishment and the 'user institution'. But it is clear that Richterich & Chancerel, writing under the ægis of the Council of Europe, have a wider interest than merely deciding what a learner wants to learn. The premise is that needs analysis is part of a social (and political) stance which allows the individual, by implementing his wants, to develop as a responsible contributor to a democratic state. The ideals are 'to facilitate the free movement of people and ideas in Europe... to make the process of language learning more democratic ... to provide a framework for close and effective international co-operation in the organisation of language learning' (Trim 1978:1).

Possible solutions

But no such broad idealism can inform, except at very long range, the procedures of the present project. Practical solutions to the problems arising from the notion of levels in devising assessments for the classroom may however appear if the ideas discussed above are considered in conjunction with the tests analysed in Chapter 2. One suggestion for example is that tasks can become progressively more complex as a reflection of ever higher levels of demand. The theory supporting this is expounded by Corder (1980) and realised in (eg) the ARELS oral (ARELS Examinations Trust nd). Another is to set increasing demands of social interaction. This is the progression suggested by Trim (1978) and put into practice in (eg) the OMLAC materials (OMLAC 1978), which are related to a day trip at Level 1 and to

a stay with a family at Level 2, an approach which was adopted by many of the GOML groups (Harrison 1982a). A third communicatively appropriate cline for levels is to set up criteria in detail in accordance with a system of recycling of functions at ever higher levels of social sophistication, which is envisaged by Wilkins (1973) and attempted in (eg) CAE (London) (ULSEB 1986), where 'has shown the ability to' statements on certificates reflect the detail of the test specifications at each level. A fourth strategy is to sidestep the levels problem altogether. If the development of assessment materials is to be based on the demands which may be made of the learners by the work done in a particular classroom, it might in theory be possible to set tasks which could be tackled at any level, with less being expected of participants at the lower levels and more sophisticated responses required at the higher. A theoretical example of this approach is suggested by Harrison (1979:12) and a practical one is the use of overlapping sets of reading texts for the three levels of CUEFL (RSA 1985). In practice however there might not be enough variety either in language demanded or in task type for communicative interactions to be developed in this way.

For this project

In the classroom, the level of the learners has already been ascertained by some kind of placement or promotion procedure, and this level is put into practice by the teaching materials, generally a course book. For assessment which is intended to reflect the course content therefore, it seems as if concern with levels in general is beyond the remit, in that the task of setting levels has already been done by the demands of administration for homogeneous classes and the responses of publishers to customer demand. In this case, the project's aim must be to reflect the levels realised in the course book (or other material in use for teaching/learning purposes), and to be guided by its organisation. Yet within this given level for the class, the assessment system still has to decide what is acceptable in a learner text and what is not. For example, a present continuous instead of a present simple might be allowed to pass at lower levels, but not at intermediate; 'past machine' could be regarded as serving for 'time machine' at intermediate, but not at advanced. Some guidelines may however appear when the approach and content of the course book is considered in detail as a model for the approach and content of the assessment: this is the province of the next chapter.

3.3 Judgements

There are several aspects of judgement which affect all testing and assessment procedures, from administrative decisions about who is to be tested, for what purpose and when, down to the finest details of a marking scheme, and all of them can cause difficulties. The two especially problematical issues for assessment which claims to be communicative are reliability and marking systems.

Reliability

The account of tasks given in the preceding section has been in essence a discussion of validity - what is the content of the assessment system, how it relates to what went before (achievement, which in the present case is classroom activities and learning) and how it relates to what comes after (proficiency, which means how well the student applies what he now has available to a new challenge in the form of a task which is related to previous work, but expects him to weave in strands from other learning). This section is more concerned with reliability - the problems of ensuring that as far as possible the outcome of an assessment is *consistent* as to marking (both between markers and between judgements made by the same marker); *replicable* as an assessment (the same material will produce parallel results on a different occasion); and *generalisable* to other uses than the present (new material can be used on the same basis to produce similar results). This last condition includes aspects of validity, since different content will be relevant for different learners.

Lado illustrates the concept of reliability with a question: 'does the test yield dependable scores in the sense that they will not fluctuate very much so that we may know that the score obtained by a student is pretty close to the score he would obtain if we gave the test again? If it does, the test is reliable' (Lado 1961:31). More recent writers equate reliability with overall consistency, a wider application of the concept, elaborated by Bachman (1990), who refers to 'classical true score measurement theory', with its references to true score, error score and observed score, and follows this with comments on its shortcomings; then deals with correlations between parallel tests; internal consistency; and intra-rater and inter-rater reliabilities. This is an impressively comprehensive coverage of reliability as traditionally understood.

But reliability is particularly difficult to establish in communicative assessment, for several reasons. The emphasis on integration rather than on individual items each testing some example of the same skill, means that it is no longer certain what credit for a single 'item' represents. Communication is considered to depend on personal interpretations - of what is said and what is to be said (and equally of what is written and what is to be written) - so that there is rarely such a neat outcome as an untrammelled right answer. As Davies observes, 'once we admit into the equation the testee's awareness of context everything becomes possible' (1988:41). One of Davies' examples, however, is a sequencing item using *farmer duckling killed the the*, and Davies suggests that the correct response 'must be' *The farmer killed the duckling*. But one of Hudson's divergent thinkers (Hudson 1967) might well argue that the duckling could kill the farmer if it crossed the road in front of his motorcycle or if it gave him fatal gastroenteritis: Davies does not carry his 'everything becomes possible' far enough. Another difficulty in demonstrating reliability in communicative assessment is that one of its aims is to offer the student

the opportunity to show how much he can do with his present equipment of language, so that the more credit he can muster the better, and if the whole class reaches a satisfactory standard (however that may be defined), congratulations are in order, rather than disappointment at the resulting skew in the distribution which makes reliability calculations otiose.

Early doubts about the value of reliability as understood in classical theory became more insistent with the discussion of communicative testing in general (Morrow 1979) and of the importance of the individual contribution to speaking tests (Underhill 1987). Both these polemicists were derided in their time, but more recently their doubts have been supported in discussions of 'a paradigm shift, from psychometrics to a broader model of educational assessment' (Gipps 1994) or 'a profound theoretical reorientation' (McNamara 1998). At the same time, and not coincidentally, there as been a move away from traditional views to a concern with social values and ethics (eg McNamara 1998).

An interest in the individual's achievement rather than comparison of candidates in groups suggests that other means of interpreting results need to be found. The notion of criterion-referenced assessment is defined by Brown as:

> Assessment that provides information about the specific knowledge and abilities of pupils through their performances on various kinds of tasks that are interpretable in terms of what the pupils know or can do, without reference to the performance of others.
>
> Brown 1980:vii

This approach apparently gives the student priority. He is not being compared with others in order to show his relative achievement, he is on his own facing and dealing with a task designed to show what he can do. Results are to be reported as indicating 'mastery' of whatever the criterion is, as represented in tasks forming a 'domain' in which the student is to show mastery. But difficult questions arise: what is the criterion? how is it represented in tasks? what cut-off point represents mastery? - if less than 100%, in what sense is it 'mastery'? The only way to answer questions like these is to refer to the performances of a group of students: the criteria and standards of a test need to be norm-referenced before it can be described as criterion-referenced. Gipps reviews the progress of this approach from 1963 on, concluding that 'strict criterion-referenced assessment is clearly unmanageable and undesirable' (1994:96).

The introduction of Item Response Theory (IRT) into the field of language testing in the 1980s (Baker 1997) provided a new basis for statistical explorations of test results. It was described by Weiss in 1983 as an area which promised to have '...profound implications for the improvement of psychological measurement and for the solution of a variety of applied problems that have not been adequately solved by over half a century of classical psychometrics' (quoted by Baker 1997). The basis of IRT is the relationship between the characteristic measured by a given set of

items, conceived of as an underlying continuum and represented by a numerical scale, and the responses of testees to suitable test items. The main assumptions of IRT are: that the item characteristic curve (ICC) produced by responses states the relationship between probability of a correct answer to an item and 'ability' level; that items measure a single ability or trait; and that 'the probability of a person answering any one item correctly is not affected by information regarding that person's success or failure on any other item(s)' (31). But even if the procedure seems to deal with the individual, it does so by setting him up against the results of group testing, so that the generalities of bulk statistics still apply. Buck (1994) maintains that all three models of IRT are 'statistical models and have no psychological or cognitive justification'.

In spite of its apparent exactitude, mathematically achieved, the procedure is bound up with probabilities, just as with classic psychometric models: we are guessing (or, as in actuarial calculations for insurance, betting; or, as in medical research, predicting) that circumstances are, or will be, as the calculations lead us to expect. Individual factual instances are not catered for (eg a stated high proportion of smokers are reported to die annually from carcinogenic diseases, but individual grandfathers may live to 94 still smoking 40 cigarettes a day). This kind of approximation is rather far from the aims of the present project, which are centred on relating assessments to previous work in class and offering the individual student and the teacher practical guidance on how far they have succeeded.

Gipps (1994) suggests that the term 'reliability' should be dropped in favour of 'comparability', which is related to the use of an assessment - 'if performance assessment is used for accountability purposes then great care will need to be taken to ensure comparability; for teacher assessment for formative purposes comparability is of lesser concern.'(171) The difficulty however is achieving that convenient (even if in current circumstances spurious) 'proof' of consistency which traditional reliability offered. Parlett & Hamilton (1972) and Parlett (1974) suggest 'illuminative evaluation' as a way forward. This is concerned with description and interpretation rather than measurement and prediction, and it operates by means of an 'information profile' which is to be derived from any of a range of resources - observation, interviews, questionnaires and documentary and background sources. This seems to be a long answer to a short question - Is the system consistent? - but the principles on which it is founded are attractive: open-ended exploration with no preformed prejudices, focused enquiries, and accurate and sensitive reporting on well-organised data (Parlett 1974).

Marking systems

The judgments most commonly associated with testing are those made in the marking of students' answers or responses, though they are in fact only a small (though vital) element the whole series of decisions which need to be made in the trajectory from test specification to results. Other decisions may have more effect

on the overall shape and approach of the assessment than the mere allocation of a value to a candidate's response, but this valuing is nevertheless a crucial point in the process. The tester's approach to it colours (or is a reflection of) his attitude to the learner. It is in this area that the dichotomy between communicative learning principles and traditional testing principles is at its clearest. Historically, 'subjective' marking tends to be associated with a negative attitude, the expectation that the candidate will not live up to the model set by some mythical perfect answer. Essays written in examination halls on whatever subject will be considered excellent if they reach 70%. The procedure is associated with a concern with 'standards' which are destined to be unmet by candidates. On the other hand, 'objective' methods (though, as Pilliner (1968:21) points out, they are objective only in the marking) may assume that 100% is a possible achievement, and this is plausible in mathematics or perhaps in the sciences. However, it is unattainable in language assessment in general (even if it may be achieved, with dubious relevance, in some 'objective' testing procedures) because of an inevitable divergence of views about what is 'correct', or as currently formulated, 'appropriate'. A positive approach can take this divergence into account by accepting what candidates offer, preferably by setting up the marking on the basis of trials and/or pretesting, as exemplified in Harrison (1975) and Swain (1985).

Marking of productive language use (writing and especially speaking, with their support materials) has been described as falling into two systems: by categories and by units (Harrison 1986). The categories system is based on concepts reflecting in some way the reasons for assessment in the first place, principally relating back by way of content validity to construct validity. Heaton (1988:99) suggests a variety of criteria, for example grammar, vocabulary, phonology, fluency. The principles of this system are implemented by allocating a proportion of points within a stated total (say 4 concepts x 5 marks = 20). Another approach is to use rating scales of the kind which list criteria and set a range of awards, say 1 to 5, for each, and the assessor's task is then to locate a candidate's performance between extremes, for example, between 'foreign' and 'native' against the criterion: style of expression (Cohen 1980:120). This is a similar approach to that of the early FSI oral (Jones 1979).

In the 'units' system of marking, units represent individually assessable entities within a test and are marked cumulatively by the addition of credits to reach a total, which may or may not have been predicted, for example awarding one mark for communication and another for correctness to succeeding transactions in a speaking test (eg Ilyin 1976, Groot & Harrison 1979). In principle, category systems include scales and band scores when used in the assessment of language production: these operate differently from the traditional essay marking only in the fact that the matching of student output is made with a globally described level of performance which incorporates several concepts at once, rather than a set of scores allocated to a series of concepts and totalled into a single number.

Category and unit systems are two routes to the same end point (a score) but the high road (categories) starts from constructs and arrives at judgement in the form of opinions, whereas the low road (units) starts from individual tasks and finds its way by discovering what appears en route, achieving understanding and interaction with events along the way. A unit system therefore seems more suitable for communicative testing, which is concerned with a developing interaction between at least two participants and the assessment of outcomes rather than artifacts. Applied to the assessment of speaking, for example, it deals with textual contributions to the discussion rather than conceptually linguistic connotations of what is said, by taking both the language used and the implications conveyed by it as explorations of meaning and the building of understanding between participants.

However, there are two major problems with the units system. One is how a unit is to be defined in continuous text. When an exchange occurs on a simple turn-taking basis, with one statement followed by another as in a simple oral test, there is no difficulty in deciding what constitutes a unit: its boundaries coincide with the beginning and end of each individual's contribution. When a more informal discussion or narrative results from a task, boundaries are difficult to set. For example, a pair of learners working together may produce exchanges in which an opinion or a story continues over several clauses (complete sentences being rare in informal speech). The problem then is for an assessor to recognise divisions between the points made by the speaker or to break up the text on the basis of grammatical form. Neither of these systems of chunking can be based on simple, clear-cut boundaries.

The other problem is level of attack. If a judgement is to be made on a 1 (for communication) + 1 (for correctness) basis, the decision is affected by how error, or in more communicative terms, appropriateness, is treated. The system is highly effective with assessments at a low level, because candidate contributions are short and error is likely to be gross, but both problems become more intractable as assessments reach higher levels, unless the exchange is tightly structured and so likely to be less 'realistic'. (These problems are discussed further in Chapter 7.)

Nevertheless, the system has direct relevance to the principles on which, it has been argued, communicative assessment stands. Its main theoretical advantages are that it directly echoes the sequence of understanding followed by judgement which occurs in language exchanges (so that in practice it is straightforward for the assessor to apply); that it is reliable in application because it is limited to the award of three rather factual elements for each contribution (which again helps the assessor in practice by restricting his possibilities, giving him fewer categories to learn and simpler decisions to make than with more detailed schemes); and that it allows for variation beyond the simple communication + correctness criteria, since candidates may make their own personal contributions: they are not expected to produce a response as predicted by a mark scheme, but may be as inventive and

original as they wish, provided the result is appropriate to the situation.

In view of these positive benefits for assessment, the intention is to apply the unit system in some form in this project, as being generally in sympathy with communicative principles.

If additive scores and the manipulation of numerical data is problematical, perhaps words can provide more directly helpful indications, as descriptions of quality rather than indices of quantity. Assessors can work with characterisations of what is expected at each level on a scale of descriptions. These descriptions are usually produced as an individual expert's - or, better, a group of experts' - view of what is a good performance and what is a poorer one, within a range of possibilities at a given level. An individual candidate's performance is then matched with a description. (Examples are given by Jones 1979, Carroll 1980, Carroll & West 1989 and an exhaustive account of scales of language proficiency is given by North 1995.) But the descriptions are not normally distilled from actual performances: they are idealistic likelihoods, so they are in effect only a little more actual than scores with suggestions for interpretation attached. For example the ELTS assessment scale as first devised (UCLES/BC nd) has been criticised as 'attempting to describe not what actually happens in communicative situations, but what communicative theorists think happens in communicative situations' (Fulcher 1987:290). But the levels do eventually derive from data, either through experience in use, or, as in North's work, by asking experts to understand and then grade given statements (North 1995:158) so as to build up a consensus on what the wordings mean. Fulcher (1996) suggests a data-based approach to scales: he reports on an analysis of transcripts from interviews which found six phenomena interrupting fluency. Chaloub-Deville (1995) suggests further a research approach 'that derives scales empirically according to the given tests and audiences', which extends the concept to include variation according to the uses to be made of results.

But even in these schemes, candidates are being fitted into pre-described behaviour patterns (however empirically established) at predestined levels, when the individual may be at various levels on the various elements included in a description - another aspect of the four-skills reporting problem reported above in relation to Morrow (1979) and CCSE (UCLES 1990). This is part of the difficulty of what is to be included at any given level: in theory, the same attributes should be included each time, but it may not be possible to find an exponent for a particular aspect of, say, phonology at all of nine levels of a scale. It is also difficult to find positive wordings for increasing levels of performance without using value words such as 'better' or 'more', which are open to varying interpretation and are somewhat less than informative. In his discussion of bands and scores, Alderson (1991) details problems and difficulties and then points out that scales can be written for and used by different interpreters - users, constructors and assessors. His summary of the advantages of reporting scores on scales suggests that scales can provide

information about what a score means; can help to reduce the spurious impression of accuracy that a score gives; can help to improve the reliability of subjectively judged ratings; and can provide guidance to test constructors (Alderson 1991:85). The chief drawback however is that, in the interests of generalisation, candidates are being labelled with concepts which are only an approximate fit to a wide range of possibly relevant achievements.

With the intention of setting up more fitting scales, Upshur & Turner (1995) developed an 'empirically derived, ordered set of binary questions relating to boundaries between levels on the performance being evaluated' (1995:3). The procedure involves teachers in deciding on the most important elements in the successful completion of a task and then arranging these elements in a structure which allows yes/no decisions to be made on responses in an order which relates directly to marking, This system is demonstrated as improving the reliability of scoring, but has two major disadvantages: that each task needs its own set of questions and that group decisions agreeing the questions require extensive discussion time.

But if descriptions are in principle acceptable, it might be possible to write statements to fit candidates rather than fit candidates into pre-existing statements on the basis of 'I have done x' rather than 'He can do x', or even 'He has shown the ability to x'. This is the principle of Records of Achievement (ROA), adopted as a policy for all secondary schools in England and Wales in 1984 by the Department of Education & Science and the Welsh Office. There are various kinds of record, but 'the main feature which they have in common is that they set out to record achievements or experiences of young people, or assessments of their personal qualities, which go beyond examination results' (DES&WO 1984:2). This was official approval of a movement which started in 1969 with the Record of Personal Achievement scheme, developed by a committee led by D R Stansbury and subsequently adopted by Wiltshire County Council and distributed by them in response to demand from schools in other parts of Britain, for example Yorkshire, Somerset, Scotland and Northern Ireland (Swales 1979). Further early examples are the Record of Personal Experience (Stansbury 1974, 1980) and the Scottish pupil profile system of 1977 (Broadfoot 1980). The principles and practice were taken up and further explored by the Schools Council (Balogh 1982) and the National Foundation for Educational Research (NFER) (Goacher & Reid 1983). A series of pilot schemes was promoted and financed by the DES and evaluated nationally (Broadfoot et al 1988 and 1991). But in the summer of 1989, possibly because it foresaw problems of control as much as finance, the Government 'turned its back entirely on records of achievement, except for the reporting of national curriculum attainment' [ie examinations] (Munby 1989), though local schemes still persist. A useful review of the whole field, with varied examples, is provided by Broadfoot (ed) (1986).

The application of statements to assessments of communicative language use seems to have advantages over other reporting systems, both verbally descriptive and

statistically derived. Words can tell us what has been achieved as an earnest of what proficiencies may be expected. But there are still enormous problems, for instead of the common currency of figures (however probabilistic and therefore approximate), the medium is words, with all their potential for disagreement and misunderstanding. As Brindley (1998) points out, in his extensive discussion of schemes for 'outcomes-based assessment', 'assessment criteria will be interpreted differently by different audiences according to their previous experience, unconscious expectations and subjective preferences regarding the relative importance of different communicative criteria'. But Gipps suggests that better models for current interests are procedures such as 'standards-referenced assessment', in which experts' (teachers') standards are encapsulated in verbal descriptions and then illustrated with exemplars which specify the standards to be applied; and target-related assessment, in which 'the targets are described in the syllabus; teachers are given explicit criteria for judging student performance; and there are descriptions of what students do at any particular stage (i.e. exemplars).' (Gipps 1994:94-5). These approaches differ from those reported on by Brindley (1998) in that they make a two-pronged attack on reliability by including both explicit criteria and exemplars of actual performances, thus approximating more nearly to Fulcher's requirement that scales should be drawn up on the basis of performances rather than speculation. In spite of the difficulties of drawing up criteria which are explicit enough, this is an approach to assessment which is most likely to be consonant with the aims of the present project.

3.4 Summary

This chapter has attempted to set out the factors which will most affect the realisation of a practical scheme for assessing students on the work they have done over a period in the classroom. Tasks need to be set up with the consideration of multiple factors beyond any realism they may have for the students concerned, such as the way in which they are presented, the relationship between quality of response and difficulty of problem (Pollitt 1991), the nature of the demands to be made on participants (Skehan 1998) and the ways in which they are organised so as to provide material for informative assessment. The issue of levels has been side-stepped with the justification that any scheme of assessment which closely follows classroom events is governed by the pre-existing allocation of learners to classes and course books to learners, but it remains to be seen whether this stance is adequate for the reporting of useful results. Judgements have been related to a discussion of reliability, but perhaps more important for the project is to find some way of arriving at a system of reporting (preferably in words rather than in figures, though exploring the possibilities of both) which will give students and teachers useful feedback on how successfully communicative the students have been. The next chapter now explores how the ideas discussed in the previous three chapters can be applied to practical ends.

Chapter 4

Analysing the course book

4.0 Introduction

In this chapter the aim is to report on setting up a systematic procedure to provide a basis for assessments which will reflect previous learning in the classroom. Reference points are needed for an analysis of what has been presented as instruction, and the simplest solution seems to be to ask the course book to provide them, since it is an important factor in the majority of classrooms. But the course book can contain only part of what is presented and discussed in any lesson, for teaching materials can only suggest what a teacher might do. As Allwright points out, teaching materials 'contribute to the management of language learning, but cannot possibly cope with many of the important decisions facing the "managers" working in their various situations' (Allwright 1981:9). Nevertheless, the course book may provide the starting point for the development of a universal system of analysis which can be applied to any classroom activity whatever its origin - for example materials provided by the school, the teacher's own materials, supplementary activity books, or work with authentic texts. The analysis is to be worked out on an empirical basis, that is, starting from what exists and building up from trials with a range of course books until a consistent form is reached which can then be applied to the course book in use in a particular classroom. It is hoped that the resulting data can provide the basis for a test specification which is demonstrably derived from immediately previous learning and which leads on to relevant assessment.

4.1 The course book

For and against

In discussion about course books, there is some confusion in nomenclature, some writers referring to 'textbooks', though this term seems more appropriate for books used in teaching such fields of study as History or Physics, and defined by Chambers (Schwatrz et al 1988) as 'containing the main principles of a subject'. Both expressions occur in the sources referred to in the following account, but the connotation is the same: a coherent series of units for learning at a given level published in one volume (or sometimes two, if there is a teacher's book). The term 'course book' is however used throughout in this thesis.

There are objections, both theoretical and practical, to the use of course books in

language classes. Those objecting are well represented by Sheldon, who describes them as a 'necessary evil'; but in response, Whitney summarises the position more positively. 'Although not every teacher uses them - indeed, some teachers, especially native speaking teachers, are frequently dismissive of them - textbooks are an inescapable part of our professional lives. ...There are in fact large numbers of teachers who, for a time at least, seem to be quite happy with the books that they are using, and for whom, in practice, the textbook is indeed necessary but also helpful' (Whitney 1988:235). Sheldon goes on to complain that they often fall short. 'ELT books are frequently seen as poor compromises between what is educationally desirable on the one hand and financially viable on the other' (Sheldon 1988:237). Allwright takes a wider perspective, including in his discussion all kinds of teaching materials, and offering two different views of their role: either to make sure the syllabus is covered, that the exercises are well thought out, and so on, which he calls a 'deficiency' view, making up for the teacher's potential shortcomings, with the result that the best teachers will not need the materials at all; and a 'difference' view, which sees two kinds of expertise at work, that of the materials writer and that of the teacher, allowing the best teachers to get on with fostering language learning in the classroom (1981:6). O'Neill, in response to Allwright, deals with what he calls 'textbooks', though they are only a subset of Allwright's 'teaching materials' (which include 'ideas books' and 'activities books'). O'Neill maintains that textbooks are needed to set a framework within which teachers and students can look ahead, look back and catch up, and in any case it is for the teacher to adapt and improvise: the textbook is only a jumping-off point for the teacher and the class and should not aim to be more than that (1982:110). Medgyes (1986) represents quite forcefully the case against the textbook: 'it is too general, boring, stuffed with cliché characters; it usually restricts activity to language presentation and controlled practice instead of stimulating real interaction. ...What is advocated as a substitute is a wide stock of flexible and authentic "supplementary" materials' (110).

Such supplementary materials have existed in parallel with the course book since before 'interactive activities' were introduced to promote communication in the classroom. The relationship between these and course books is partly complementary and partly competitive. No single course book, whatever its authors claim, can be comprehensive enough to meet all demands when one of the principles of communicative approaches is that variation is essential to meet local needs. Supplementary material is therefore a valuable, if not indispensable, resource. On the other hand, communicative tenets may be unacceptable in many cultures where seniority of age and experience are inbuilt and not to be easily overturned by radical thinking among linguistics experts and English language teachers. The problems are discussed by several commentators, for example Maley (1986), who compares current practice with communicative approaches and finds the change from one to the other problematical; Edge (1987), who points out that

communicative methods may contradict educational cultures and may also suggest ways of behaving which are proscribed in students' everyday lives; and Jones (1987), who suggests that the majority of ELT practitioners world-wide will either not receive, or will misinterpret, the [communicative] message. The course book authors' answer to the potential puzzlement of teachers seems to be to offer more and more plurality in syllabuses and more and more advice on how to use them. But there are other unavoidable conflicts: Hutchinson & Torres (1994) point out the mismatch between moves towards greater negotiation and choice for learners on the one hand and the development of ever more structured textbooks on the other. They admit however that this structuring is essential, for psychological and managerial reasons and for security in the face of change.

In spite of all its defects the course book remains the main driving force behind the work in most language classrooms, setting overall aims and objectives and at the same time detailed paradigms for teaching methods, so that they seem rather inevitably to provide material for analysis. They cannot offer patterns and instruction for all that happens in even the most devotedly controlled classrooms; indeed, in the hands of skilled, imaginative teachers, may represent only a skeleton on which to build a learning sequence. But for present purposes, which require a substantial and stable body of material on which to build a system of analysis, the course book seems the most economical and realistic starting point.

Authors' introductions

If the course book is to be the basis for analysis, it might be informative to consult course book authors to investigate what they consider to be the important aspects of their work. They tend to take for granted that anyone who is reading their Introduction is already committed to the use of a course book and needs persuading only that theirs is the one to choose. A selective survey was undertaken of 17 course book introductions dating from 1975 to 1996, and these represent at first sight some variety in authors' promises.

The earliest course books claiming to be 'communicative' (Abbs et al 1975, O'Neill & Snow 1977, Alexander 1978) are intended to represent a radical departure from a 'traditional' approach, eg: 'very different from that of the traditional structurally graded textbook' (Abbs et al 1975 *Strategies*); 'pioneered the new communicative approach' (O'Neill & Snow 1989 *Crescent 9A*); 'radical departure from traditional course design' (Alexander 1978 *Mainline Beginners*). After a few years, authors begin to emphasise a variety of content going beyond the functional-notional, in apparent reaction to the idea - never seriously proposed by any of their earlier colleagues - that functions alone are sufficient basis for learning: 'organisation is... based on a combination of eight different syllabus-inventories... systematic teaching of vocabulary... dictionary use' (Swan & Walter 1985 *Cambridge English course*). Another apparently reactionary move is a 'return' to grammar - a renewal which seems more polemical than actual: 'comprehensive coverage of the English

grammatical system... grammar areas are dealt with in more depth than is usually found in course books' (Soars & Soars 1986 *Headway Intermediate*). A more radical development is to call on the findings of a corpus to indicate what English is actually used by L1 speakers: 'a mass of information from the careful examination of many millions of words... boldly features spontaneous conversation, language at its most natural' (Sinclair, 'Foreword' to *Collins Cobuild English course* Willis & Willis 1989). Later course books are still attempting to combine the old and the new so as to appeal to as wide as possible a range of users: 'alternating units on grammar and vocabulary... highly original tasks... innovative Teacher's Book material... course can be tailored closely to the needs of individual classes' Doff & Jones 1994 *Language in use*); 'combines the best of traditional methods with more recent approaches, to help students use English both accurately and fluently' (Soars & Soars 1996 *New Headway English course*).

There is no justification in any of these expositions for the use of a course book in principle rather than the use of the course book in question. The only exception to this rule which has been discovered is in the 'Key-Notes' paragraphs in Soars & Soars' *Headway Advanced*, where the authors suggest that it is only recently that course books have been written for the advanced student, and so find the need to explain why they have produced one. Even these students have gaps in their knowledge and 'it is possible to devise a broad-based syllabus consisting of revision and new input of both grammar and lexis. Added to this, there is a strong reassuring feeling for a student who can say, "This is my English course book; I know what it's doing; I can learn from this," and there is no reason why this should not also be true for the advanced student' (Soars & Soars 1989:i). Further on, they add: 'Just as students benefit from having a course book, so teachers should feel reassured by having a framework and a direction on which they can build' (ii). These comments could clearly apply to students and teachers at all levels, and the fact that they are made here, in the introduction to an advanced course, may reflect some unmentioned criticism of course books in general (or of one at Advanced level) which the writers feel they ought to refute.

Two interesting points do however emerge from the review of course books: firstly that the writers almost entirely ignore the proposition that the teaching of grammar is not necessary for learning by communicative principles. Some over-enthusiastic teachers (perhaps misinterpreting Krashen's views on language acquisition) seemed to be claiming in the 70s that grammar was acquired along the way without direct instruction, eg O'Neill & Snow's suggestion: 'The communicative approach focuses on acquiring the grammatical rules by using the language' (*Crescent*, 1989:23), though they then add, 'However, there are times when the rules need to be explained'. Of the 17 texts consulted, only *Crescent* and *Communicate* (Morrow & Johnson 1979) omit grammar teaching altogether, and *Quartet* (Grellet et al 1982) makes it optional.

Secondly, the lack of variety between the earlier and later books is remarkable. For example, the arguments for *English Alive* 2 (Nicholls et al 1978) and *Approaches* (Johnson & Morrow 1979), are improved on surprisingly little by *New Headway* (Soars & Soars 1996), published 17 or 18 years later. Regarded as input for the analysis of course books, the survey shows only that the standard elements in a course need to be included. 'We should always remember that new ideas do not supersede earlier ones. Rather, they coexist with them, so that we can see older ideas in a new light' (Alexander 1978). But there is one glimmer of guidance: through the list there is a discernible increase in attention to the learner, who in 1975 is provided with skills (Abbs et al 1975) but in 1986 is expected to take responsibility for his own learning (Soars & Soars *Headway Intermediate*) and by 1993 is to be subjected to a systematic learner-training programme (Radley & Millerchip *Workout*). This is encouraging for a learner-oriented approach, as in the present instance.

Course book evaluation

Another possible source of points of interest for setting up an analysis which may lead to later assessment is the evaluation of course books: is there a framework or paradigm which, intended to formalise the consideration of course book quality, could help to generate an assessment syllabus? Sheldon (1987), in his introduction to a 'collection of essays' about materials evaluation, suggests that too many course books are vague about target learners, explain grammar too technically, are unsuitable in layout and rarely explain their rationales (p3). Cunningsworth (1984:74ff) offers a long checklist of evaluation criteria, and though they are set out in the main as detailed questions for teachers to ask about the content and approach of a given course book, the answers say little about the demands made on learners by the language tasks set. McDonough & Shaw (1993) propose two approaches to evaluation, external and internal. The former is very general, dealing with blurbs and writers' claims, but 'internal' evaluation promises an 'in-depth investigation' into the materials. The results are disappointing. The topics on which questions are to be asked include presentation of skills, grading, text beyond the sentence and authenticity, but go no further than these shallows: what students actually have to do to engage with the material offered remains unexplored at the deep end.

One of Sheldon's 'collection of essays' however is a contribution by Breen & Candlin, who suggest a series of questions for teachers which take a positive view of what might be found in a course book. Of the 34 questions they offer, the following three seem significant:

> 2. [when they finish their course] What should [the learners] be able to do in and with the language?
>
> 8. What procedure or sequence of work does the learner have to follow in order to be successful at the task?
>
> 23. In what ways do the materials involve your learners' values, attitudes and feelings?
>
> Breen & Candlin 1987:14-20

These questions suggest directions in which the development of the present analysis might run, and will be will be one of the bases for the discussion which follows.

Summary

The justification for using course books as the basis for this present analysis is that they are available in any classroom, are common to large numbers of classes (as teacher materials are not) and are relatively static (as teacher responses to the classroom dynamic cannot be) and so provide relatively easy access. The categories eventually arrived at can be applicable whatever the book used. An extension can even be envisaged in which the system is applied to classroom events as they occur, so that when tangents suddenly appear and are pursued, or the activity proposed by the course book is abandoned and a new activity is substituted, these spontaneous events can be written up and included in the subsequent assessment. But the analysis of what happens in the classroom needs to start not just from the content of the course but with the students' reaction to it. The important questions are: What are the learners being asked to do? and What does the activity look like from their point of view? The answers to these questions can then guide the design of assessment material which aims to cover the same ground as the learning.

4.2 A system for analysis

Events for the learner

Little of the above exploration of guidelines to course book analysis for teachers is of much help. What is needed for the present project is detailed information about what has happened in the classroom against the background of the course book, which, as noted above, cannot cover all possibilities. The aim is to develop an analysis system which will enable the learning activities to be paralleled by assessment materials in which, as far as possible, the same things are planned to happen. Perhaps, taking further the questions asked by Breen & Candlin (quoted above), an analysis could be usefully based on the viewpoint of the learner. Breen has further helpful suggestions to make in this direction. In his discussion of how learners recreate a workplan he includes their (the learners') questions: 'Why do I need to do this?', 'What am I working on?' 'How do I go about it?' and 'Where am I working?' (Breen 1987:23-46). This approach seems to offer a useful perspective. If the analysis looked at the course book from the learner's standpoint it would bring into focus what was actually being demanded by the material offered.

Three examples of analysis systems drawn up to investigate examinations in modern languages provide some general guidelines for this approach. James & Rouve (1973) described language performance by undertaking a survey of syllabuses and schemes of language examinations in schools and in Further and Higher Education in order 'to answer the question which we imagine would be put by any potential employer (or admissions officer) ... What can this applicant actually do?' (1973:1).

A later survey, of modern language examinations for secondary school pupils at the age of 16+, was 'an attempt to provide a critical analysis', but took on a polemical tone, attacking the examining boards for their traditional approach and their unwillingness to provide detailed information about aspects of the setting, content and marking of their examinations (Moys et al 1980). Finally, an analysis of the examinations of one board, the Institute of Linguists (Harrison 1979, 1980a) developed a system for comparing the Institute's syllabuses with those of other examining boards, starting from the question: 'What does the candidate have to do in order to give the examiner something to assess?' In this analysis, a coding was applied to the skills required for each activity, including six beyond the standard four (listen, speak, read, write): understanding of the visual, summarising, showing knowledge of content, giving a mechanical response (eg in multiple choice testing), performing some physical act, producing a non-linguistic response on paper (eg draw). This coding system was later adapted for use as guidelines for test specifications (Harrison 1983b) and becomes the starting point for parts of the coding used in the present analysis.

In view of these precedents, and because a learner-centred approach promised to be particularly apt for course book material claiming to be communicative, it was decided that the the analysis was to begin with the learner. The analyser needs to put himself in the learner's shoes: 'If I were a learner faced with this task to fulfil with this material, what would I actually *do* in attempting to carry it out?' The immediate implication of this is that the material has to be considered as a series of things to be done (ie tasks), each of which makes different demands on the learner, but together with others in the series operates within an overall structure of constant components. The first step is to draw up a list of decision points for the learner - the facts which confront him as he works through a task - in the order in which he meets them as he responds. The aim is to arrive at a logical sequence in the events which universally occur with any communicative exchange, including suppositions about their impact on the learner's mind and about his reaction to them. These ideas were elaborated in a working paper (WP1) which charted the course of the development, but it is too prolix to be reproduced entire. The following discussion is therefore based on quotations from it, suitably edited.

Categories for analysis

Several draft lists of categories were made and gradually refined through applying them to tasks set in a popular course book (*Cambridge English Course* 2 - Swan & Walter 1985). This was chosen as a starting point because it had appeared consistently over several months in the list of top selling EFL books compiled by the Bournemouth English Bookshop and published in the *EFL Gazette*.

The first workable version consisted of the following categories:

- context
  - who/where...
  - culture
  - group
- content
  - receptive
  - processing
  - productive
  - simultaneity
- language
  - structure
  - lexis
  - phonology
  - discourse

The subheadings, for example, under context - who/where; culture; group, were subjected to a detailed discussion of alternatives and possibilities, and this process was repeated for all the categories and subheadings.

4.3 Analysis Sheet 1

Layout

After further debate in WP1, an analysis sheet was designed as a form on which categories of learner decision were arranged horizontally in the sequence in which it was inferred he would meet them, with vertical space under each category heading for the analyser to enter a record of each task and its demands. *Figure 4.1* shows this early draft version. After ten or so of these forms had been filled in (amounting to some 900 entries overall), including some for *Cobuild Intermediate* (Willis & Willis 1989) - a course based on 'real English' - preliminary rules for completion were laid down to guide the analyser, both to remind him of decisions made earlier and to promote consistency. But the categories were found to be imprecise, and after several sets of detailed amendments had been made to category labels and groupings, the sheet was radically revised.

If the sequence in which the learner had to make decisions was to be followed through logically, there occurred between *receptive* and *productive* an activity labelled *processing* in the draft analysis sheet (see *Figure 4.1*) which represented a potentially vast range of mental activity. The main difficulty was categorising relevant aspects of the notion 'processing' and recording them on the analysis sheet. It had been recognised from the start that the record would have to be a highly condensed version of what appeared in the course book, but the *processing* heading represented so much more than just a link between skills (eg using reading as a basis for speaking) that it was allocated to a new analysis sheet of its own, labelled AS2, with the original now designated AS1 (The final version of AS1 will appear in *Figure 4.2*, when the development of rules for completing it has been discussed.)

Figure 4.1: Analysis sheet, draft version

ex	context	culture	group	content				language demands			
				receptive	processing	productive	simultaneity	structure	lexis	phonology	discourse

The implementation of AS1 still presented problems, however. The most difficult were the realisation of entries under 'language demands' (the last four columns in *Figure 4.1*) and the definition of the end of one task and the beginning of another.

Language content

Even taking into account the need for sampling, and hence for summarising into entries on AS1, there was an acute problem in consistent and at the same time representative condensation, especially under the headings *structure* and *lexis*. For example, the structures needed for the completion of a task could go far beyond those purported to be the learning point, and the vocabulary for one task alone could sometimes, if comprehensiveness were to be aspired to, fill the whole column on the analysis sheet. Some kind of restriction was clearly necessary, but on principles which could be implemented with consistency. A solution was found in a change of concept from the original 'language demands' to 'language essentials' (compare *Figures 4.1* and *4.2*), which helped to dilute the condensation problem: if the entries consisted only of those elements which were essential for the completion of the task, all others could be ignored. Thus to control the entries required for vocabulary, the question from the learner's point of view is 'What words are essential for my completion of this task?', and what is to be entered on the analysis sheet is a 'sample of the minimum few words without which it would be impossible for me to do the task.' (These wordings are taken from the Analysis Guide which was written in due course to explain how the procedure was to be implemented.)

Task boundaries

It became clear after the completion of about 20 examples of AS1 that there was a problem for an analyser in deciding where a task was to be recorded as starting and where it could be considered to be complete. The course book is not itself necessarily a sure guide: it may set out steps within the Unit of the course, such as 'Pre-listening task... Listening for gist... Comprehension check... What do you think?' (*Headway Intermediate* pp12/13, p6), but these represent different exercises on the same topic, whereas it is probably more important from the learner's point of view to take into account a change of focus, either from pair to group, or from one use of the text to another, for example from listening for information to discussion of ideas. Systems of classroom observation (Allwright 1988, Chaudron 1988) are of little help because they are mainly concerned with teacher/pupil interactions rather

than a learner's approach to materials, and in many cases rely on arbitrary timed intervals in a continuous flow of classroom activities.

A further difficulty may be caused by the course book writer's use of rubric, a term used here for all aspects of the instructions given to the learner as to what he is expected to do in the task. These instructions sometimes take the form of introductions, such as a mind-focusing activity which appears in the teacher's book but not in the learner's, or a pre-listening task which is then the basis for class discussion. The question then is: when is a rubric to be considered merely as an instruction for carrying out a task, and when is it to be entered on the analysis sheet as a task in its own right, and therefore constituting a new start?

The problem of defining task starts became even more obvious when, as a test of analyser reliability, a repeat analysis was undertaken of the same pages in one of the course books, and discrepancies appeared. This discussion took place in another working paper (WP2), of which extracts now follow.

The questions to be asked are the student's, eg: 'At what point does it seem to me that I have finished this task? What indicates to me that am I starting another?' It might be instructive to explore whether this approach could be useful for setting up defining questions for analysers, using some of the analysis headings as a guide.

In the *HI* (*Headway Intermediate*) example ('Pre-listening task... Listening for gist... Comprehension check... What do you think?'), a sequence of activities is suggested in the teacher's book (p 6) for the section headed 'Listening'* in the student's book (pp12-13). The starting point is

a the teacher asks the students as a class three questions given in the teacher's book; then

b the students listen to and at the same time read in the student's book the beginning of a radio programme;

c they do a 'pre-listening task' (questions to discuss as a class);

d they listen to the whole recording ('listening for gist') and discuss in pairs whether their ideas appeared in it;

e the recording is played again ("if necessary"*) and the students do a comprehension check, consisting of five factual questions about what has been said on the recording (how they are to be answered is not specified in the teacher's book - presumably as a class?);

f then "students could ask each other questions about the tape";

g "Do the *What do you think?* questions" - three general discussion questions about the topic.

* Note: 'single quotes' indicate extracts from the student's book, "double quotes" extracts from the teacher's book.

Here are seven activities each of which could be entered in the analysis as a separate task.

The discussion in WP2 explored whether both beginnings and ends of tasks should be investigated to provide rules for analysers, but it was concluded that establishing the necessary conditions for beginnings automatically solved problems with ends. It was decided that these rules for analysers should take the form of choices which would be set up as 'if/then' alternatives, and this was achieved by reviewing the tasks already represented on the analysis sheet and asking what conditions had suggested that a new task was beginning, according to the information given to teachers and learners by the course book.

The seven activities listed as *a* - *g* at the beginning of the exercise (see above) were therefore worked through in detail so as to set up the basis for a trial with course books of a limited part of the analysis sheet to help sort out possible rules for beginnings. The question at that stage was: Does a change within in category mean a new task? The WP listed possibilities:

yes, change under this heading = new task.....but only when

1 [first on new sheet] .....not explanatory or short rubric

2 topic .....beginning of new topic

3 protagonist ('who') .....'where' also changes

4 place ('where').....'who' also changes

5 'group' .....[unconditional]

6 status ('discourse') .....social group also changes

7 social group ('discourse') .....status also changes

The number of changes is not by itself enough to establish new task starts; and changes in 'skills' alone are irrelevant. In practice, the beginning of a new analysis sheet requires a full set of entries so as to establish a starting point for changes, so the first rule relates only to rubric type.

This and pairings of the overlapping 3 & 4 and 6 & 7 reduce the questions to five:

1 Is it a preparatory rubric?

2 Is it a new topic?

3 Do both 'who' and 'where' change?

4 Does 'group' change?

5 Do both parts of 'discourse' change?

A 'yes' answer to any of these questions means that the analyser is to start a new task on the Analysis sheet.

But in the process of arguing through these questions, other uncertainties appeared which needed clarification, and these were summarised as four points:

- whether size of activity is a factor
- when to include rubrics
- when to include optional activities
- whether to include function

The next stage was to apply the conclusions listed above with an Analysis sheet adapted to include only the categories under discussion, firstly to find out whether the five questions were successful in defining task beginnings, and secondly to explore how far the four points were helpful in defining the nature of tasks for the analysis.

This restricted form of AS1 was therefore drawn up and applied to four varied course books. Two of them were eminent in best-seller listings in *EFL Gazette* (*Headway Intermediate* and *Cambridge English course 2*), another was uniquely corpus-based, claiming to represent actual communications (*Collins Cobuild English course 2*), and the fourth was included at this point as an example of a course aiming at a widespread but specific audience of third-world teachers and students (*English for French-speaking Africa 2* - Mills et al 1986), in an attempt to ensure that the system could cope with other courses than those of British origin intended mainly for use by European-oriented teachers. After a dozen applications (400+ entries), the Analysis sheets for each of the books were summarised to show how often changes in task beginnings were dependent on changes in category headings.

Rubric

The most complex problem is how to cope with rubric. The distinction made above between explanatory and preparatory rubrics and the rule that activities not included in the student's books should be omitted from the analysis both turn out to be helpful, but not conclusive. The following are examples of difficulties and exceptions:

- The author's labelling of an activity as a 'pre-reading' task does not necessarily mean that the whole of it can be classified as 'rubric'. For example, in *Hl, h* was recorded as a separate entry from *g* because it included scanning of the text rather than merely discussion of points as preparation for reading.

- The fact that an activity does not appear in the student's book (even though it is in the teacher's) overrides the distinction between preparation and explanation by which preparation is to be recorded but not explanation. The exception to this rule is that when the activity is by nature teacher-activated, it should be recorded as an entry, for example dictation (*EFSA h*), the inclusion of which in the student's book would be counterproductive (eg 'Your teacher will now give you a dictation based on lines NN of dialogue X'). Another exception originally made for *Cobuild d* ("Before listening to the

tape initiate a discussion about what kind of man the writer is...", not mentioned in student's book but instigating a task in the terms laid down) is therefore omitted. This exception is the only one of its kind in the 54 entries, and though it may be plausible, must be rejected in the interests of simplicity of ruling and hence consistency. The final ruling is therefore that if an activity does not appear in the student's book it is not to be included in the analysis as a task, unless it is provided for teacher use as an integral part of the task itself.

Progressive rulings

Near the beginning of the restricted analysis exercise, several cases of amalgamation occurred. This was when an entry was made and then, in the light of evolving rulings, combined with its neighbour when the changes recorded were judged too minor to represent a new task for the learner. This development was typical of the trialling procedure, since experience of the analysis altered decisions as the rules became more apparent. Hence there were no amalgamations in the last three sheets, and unrecorded omissions were progressively more frequent. This does not invalidate the procedure, but does mean that there were a higher proportion of 'legal' entries towards the end.

The questions for analysers to consider can now be reduced to:

1 Is the activity in the student's book?

2 If it is a rubric, is it preparatory (rather than explanatory)?

3 Is it a new topic?

4 Do both 'who' and 'where' change?

5 Does 'group' change?

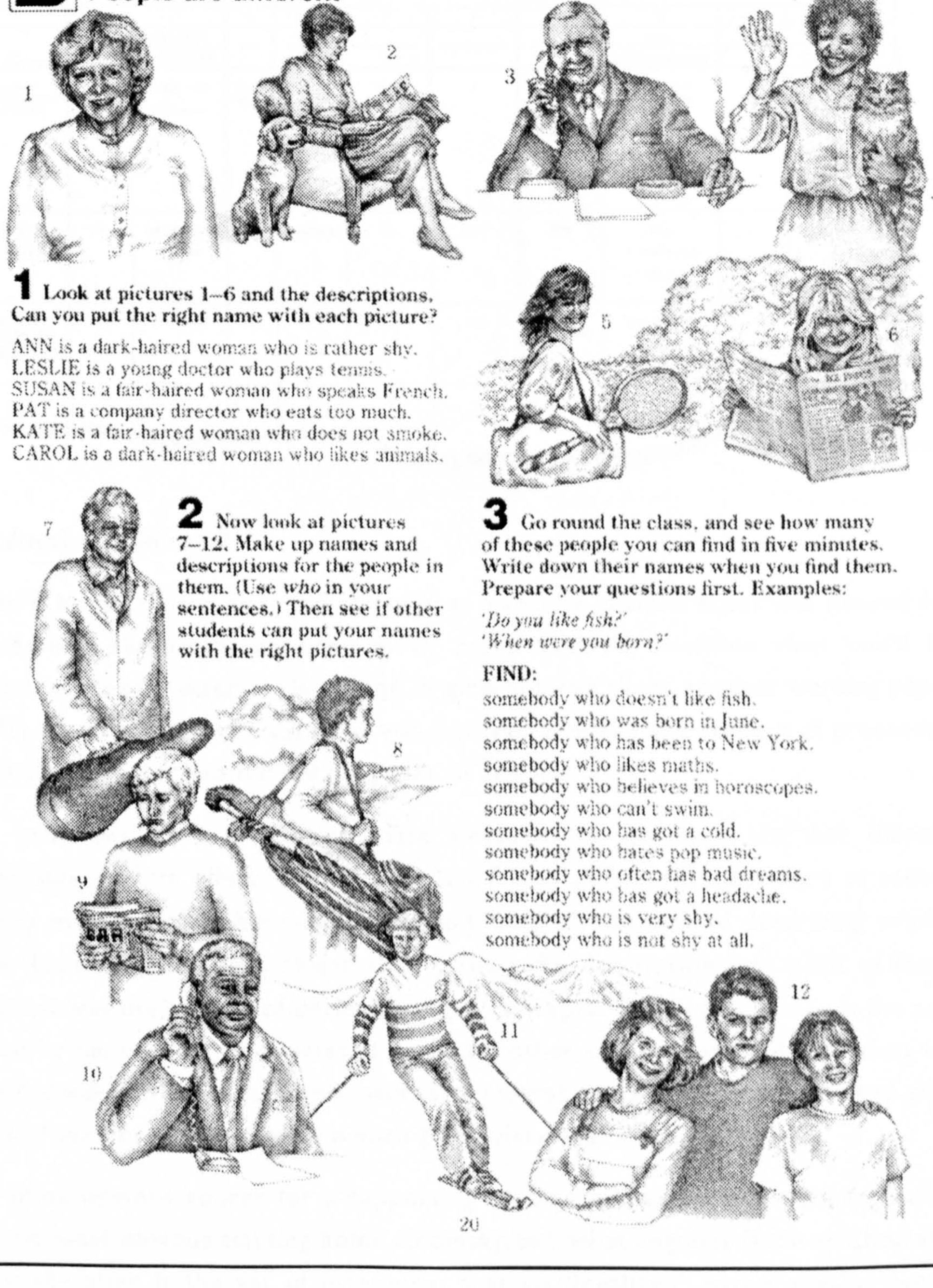
A 'yes' to the first two means that the activity is to be included in the analysis; a 'no', that it is to be omitted. A 'yes' to any one of the other three questions, taken in order, indicates that the activity is to be entered on the analysis sheet as a new task.

The final version of AS1 is given in *Figure 4.2*, which includes a sample of the analysis carried out for *CEC2*, Unit 4B ('People are different') and the relevant page of the learner's book.

Figure 4.2: sample of AS1 applied: *CEC* unit 4B (student's book)

Unit 4

B People are different



1 Look at pictures 1–6 and the descriptions. Can you put the right name with each picture?

ANN is a dark-haired woman who is rather shy.
LESLIE is a young doctor who plays tennis.
SUSAN is a fair-haired woman who speaks French.
PAT is a company director who eats too much.
KATE is a fair-haired woman who does not smoke.
CAROL is a dark-haired woman who likes animals.

2 Now look at pictures 7–12. Make up names and descriptions for the people in them. (Use *who* in your sentences.) Then see if other students can put your names with the right pictures.

3 Go round the class, and see how many of these people you can find in five minutes. Write down their names when you find them. Prepare your questions first. Examples:

'Do you like fish?'
'When were you born?'

FIND:
somebody who doesn't like fish.
somebody who was born in June.
somebody who has been to New York.
somebody who likes maths.
somebody who believes in horoscopes.
somebody who can't swim.
somebody who has got a cold.
somebody who hates pop music.
somebody who often has bad dreams.
somebody who has got a headache.
somebody who is very shy.
somebody who is not shy at all.

Swan & Walter, *Cambridge English course 2 (CUP 1985)*

Figure 4.2: sample of AS1 applied: *CEC* unit 4B

ref: CEC2, unit 4, page 20

LAC Analysis sheet 1

task	context				skills		language essentials			processing ⇨
1	2	3	4	5	6	7	8	9	10	11
4B People are different	topic	situation who/where	set	culture	receptive	productive	structure	lexis field/vocab/func	phonology	discourse relat/ass/123
(a) match (descriptions of people w pics	descriptions of people	author - self - classroom	lone/ group	Brit	vR(L)	S	rel clause (who) pres simple neg pres (does not)	appearance & characteristics of people - dark-/fair-haired young shy tennis French director to smoke - describing	-	sub - descriptions - 1
(b) invent (descriptions for pics)	ditto	self - partners - classroom	group	W European	vL	(W)S	ditto	ditto - ditto + own - ditto	-	equal - ditto - ditto
(c) find out (characteristics of people in class)	ditto	self-students - classroom	class	ditto	(R)L	WS	ques pres simple	characteristics of people	-	ditto - ditto - ditto

4.4 Analysis Sheet 2

When it became clear, as noted above, that a second analysis sheet was needed for 'processing' because it required much more detailed attention than could be accommodated between 'skills' on the original analysis sheet, another working paper (WP3) was written to discuss what was involved and how the content of processing might be characterised for the purposes of analysis.

The problems with processing arise because its workings are not directly observable and its effects not directly attributable. Though some sort of mental activity must occur in between 'input' and 'output', the task of describing exactly what this activity consists of and then turning the description into a list of found facts involves the reifying of decidedly vague concepts. For example, how is the part played by memory to be separated out from other components and how then will it be recorded in practical terms (words, numbers) on an analysis sheet? And what kinds of memory are involved in each particular case?

The most obvious source for a supporting theory seemed to be psycholinguistics, and the most obvious starting point, Chomsky, but 'what linguists [such as Chomsky] are really after is the set of utterances that an "idealized" native speaker would accept as being grammatical.' (Greene 1972: 13,27). The present effort is directed towards describing the actual, not the ideal. Other approaches were investigated, such as Garman's concern with 'such questions as "How does a speaker go about putting ideas into forms that can be expressed as patterns of articulatory, or manual, movements?"' (Garman 1990: xiii) and Færch & Kasper, who present a series of papers on introspection in second language research (1987). None of these however provided practical starting points for the kind of analysis required in the present case. But the 'processing' element does exist as an essential part of language

use, and therefore as something which needs to be recorded in some relevant way on an Analysis sheet.

Perhaps a better attack is to seek some structure within which labels can be allocated - that is, to concentrate first on how to formulate categories and then, as a result of experimental applications, consider what specifically linguistic elements should be allocated to them. Sternberg (1985) suggests that there are two kinds of theory for describing human intelligence: explicit, which sets up a detailed specification of mental structures and processes; and implicit, which relies on the perceptions of what the concept 'intelligent person' means to people in general. If the first approach is taken towards the definition of 'language processing', it should be possible to start from a set of reasonable suppositions about what 'processing' might consist of, and then test them against examples found in the course books which are the current source of information for the analysis.

In the context of making decisions in second language acquisition about task difficulty, and hence sequencing, Skehan (1992), building on suggestions made by Candlin (1987:18-20), proposes a scheme of elements representing clines along which difficulty for the learner may lie, under three headings, Code Complexity, Communicative Stress and Cognitive Complexity, with two subheadings under this last: Cognitive Processing and Cognitive Familiarity. For present purposes, the first main heading, Code Complexity, can be assumed to be covered by the final four columns of the analysis sheet, under 'language essentials'.

Skehan's elaboration under the remaining three headings (1992:198) were then listed in the Working Paper. As an example of this listing, the sub-headings for 'Communicative stress' are given in *Figure 4.3*.

Figure 4.3: Elements in communicative stress

Communicative stress

- time limits and time pressure
- speed of presentation
- number of participants
- length of texts used
- type of response
- opportunities to control interaction

(continued)...

Skehan 1992:198

Concepts of this kind seemed potentially fruitful as a basis for the present analysis, since they could be related directly to the 'What am I doing here?' perspective (glossed as both:'What's the point of my involvement?' - a question of communication; and 'What is expected of me at this stage?' - a question of resources).

It was decided to set up a checklist on the basis of Skehan's categorisation for analysis in the 'processing' column and use it on an experimental basis to see whether it would help to establish a set of categories under which judgements could be made about 'processing' for the analysis of course books. The elements in the checklist were labelled *pressure*, *cognitive processing*, and *cognitive familiarity*. But considering the point of view of the learner, as before, it seemed necessary to include the *response* required by the task, so this was added as a fourth element. These were the elements included in the first version of the checklist.

The procedure followed on from entries made on an AS1, taking the tasks listed there and continuing horizontally on to a second page, AS2. This contained a further series of columns headed with the elements which were now suggested for the proposed aspects of processing. In this way, the information already recorded on AS1 was used as the source of answers to the questions brought up by AS2. This relationship will be further illustrated in the account of the development of the analysis of processing below.

It soon became apparent that this approach needed a different system of recording: not of essential facts as in AS1, but of judgements about clines on a scale of difficulty. An arbitrary scale of 1 (low) to 7 (high) was postulated as a starting point, and events as found in the course book material were then graded against this scale for each of the four subheadings in the 'processing' column, as listed above. A review after 14 Analysis sheets had been completed, including some repeat analyses of the same course book units, showed considerable variation in gradings, implying that more stringent rules of application were needed. The ratings were not falling into patterns, as had been hoped, and there were no anchor points to which decisions on grading could be related. Judgements evidently needed to be better specified. The checklist was therefore annotated to show details of factors to be considered under the headings and dimensions of demand for each, becoming version 2, an example from which is given in *Figure 4.4*.

This further definition would, it was hoped, make subsequent application easier as well as more accurate. It became clear at the same time however that the information provided so far for AS2 was inadequate for review purposes without some indication of the reasons why the gradings had been allocated, so brief notes of one or two words were added to each entry from then on. And as a final contribution to consistency at this stage, grading was to be based on the system which starts from an assumption of 'average' unless indications are otherwise. (This approach has long-standing precedents in the oral examinations of the English Speaking Board (Burniston 1968) and the practical tests of the Associated Board of

the Royal Schools of Music). In the present case, a grading of 4 was to be regarded as an unremarked average, with comments thus being made only for positive or negative variations.

Figure 4.4: Processing: Checklist version 2

heading	*includes...*	*level of demand (7 high > 1 low)*
PRESSURE		
time available	speed of input (eg L1 speaker(s))	fast > slow
	speed required for (& in) response	quick > relaxed
N elements	either series or parallel	many > few
	participants	many > few
psychological	learner to peers/lone	stand up > hide
	competition	intense > weak
COGNITIVE PROCESSING		
creativity	finding something to say, constructing ideas, originality	required > not
matching	same information, different modes (eg spoken or written)	distant > close
	same mode, different information (eg compare experiences)	distant > close
N elements	to be processed at once	many > few
remembering	how much needed	a lot > a little
transfer	concrete to abstract	complex > simple
	abstract to concrete	complex > simple

(continued) ...

After Analysis sheet 29, an accumulation of 108 entries, a summary was made of the notes which had been written to indicate the various aspects of difficulty within the four main headings. Examples are given in *Figure 4.5*.

Figure 4.5: AS2: draft notes on difficulty levels

Pressure, grade 5:

remember, speed, invent, search, L speed, S to peers, compete, agree...;

Cognitive Processing, grade 6:

search & match, quantity material, argue, match (socio), creativity, sort essentials, explain, create, elements, matching, invent...;

(continued)...

Figure 4.6 shows how the notes were gathered under group headings and the gradings distributed across the postulated range of 7 to 1.

Figure 4.6: AS2: samples of notes on gradings

	7	5	3	1
pressure	speed L1 accent competition	remember invent S to class speed L1 discuss group	own time elements	
cognitive processing	concentrate argue pick out	remember invent match skills search infer	revise practise transfer already match	

(continued)..

The result was rather unsatisfactory in several ways, but further discussion showed how these apparent shortcomings could be turned to advantage. The exploratory method of setting up 'codings' as succinctly as possible in words or phrases to characterise each concept as it appeared, inevitably resulted in ambiguities and duplications. For example *invent* appears under 'Pressure' (column 3) and twice under 'Cognitive Processing' (columns 2 & 3); and *elements* appears under 'Pressure' (column 4) and also under 'Cognitive processing' (column 2). In the first instance, synonyms could be used to separate two applications of the same word in different areas, eg *invent* remained as an aspect of creativity under Cognitive Processing, but was subsumed into *speed of response* for Pressure. In the second case, *elements* could be graded as either following each other in a series (under Pressure) or as occurring all at once (under Cognitive Processing), so that the context in which the words were used was also specified.

For five areas of difficulty such as these, further thought suggested directions in which positive action could be taken, and these led on to another revised version of the checklist (version 3). This now included, for each area considered relevant to processing:

- a definition of the main heading, with an indication of origin and a key question the view of the learner;
- sub-headings representing aspects of the main heading;
- keywords (in **bold type**) for use on the analysis sheet to represent what is to be graded; and

• the dimensions along which the grading is to be applied

Figure 4.7 gives a sample of this layout.

Figure 4.7: Processing: Checklist version 3

PRESSURE (externally imposed; Q: how hard is the learner driven?)		
P1 time available for learner	• speed of input (eg L1)	fast > slow
	• speed of response required	quick > relaxed
P2 N elements in input	• facts (diachronous) x+y+z	many > few
	• participants	many > few
P3 psychological	• S to class	stand up > hide
	• competition	intense > irrelevant
COGNITIVE PROCESSING (internally operating; Q: what mental gymnastics are required?)		
CP1 creativity	• invent	required > not
	• structure of ideas	to find > given
CP2 matching	• same information, with different mode (eg S to W)	distant > close
	• same mode, with different info (eg compare experience)	distant > close
CP3 N elements in task	• to be processed together (synchronous)	many > few
CP4 remembering	• how much required	a lot > a little
CP5 transfer	• concrete (words) to abstract (ideas)	complex > simple
		• abstract
	(notions)to concrete (situations)	complex > simple

(continued)...

Version 3 was tried out on the course book sample used for the 'task beginnings' investigation as described above, bringing the total of entries up to 1242, which was considered to be an adequate sample for finding out how successfully the keywords could be applied in practice and to provide examples for the guidance of future analysers. Still further problems appeared, however. There were clearly too many keywords for practical application, but the evidence which had now accumulated gave empirical grounds for review and amalgamation of categories. In addition, there was still no structure for deciding what gradings should be applied within the area represented by each keyword. For example, the clines suggested under the heading 'level of demand', such as *fast* > *slow*, or *many* > *few*, did not give guidance on how fast was 'fast' or how many was 'many': some kind of calibration was needed to promote consistency of judgement.

Both the multiplicity of key-words and the definition of levels were attacked by using what Ebel calls 'global quality scaling'; more specifically, a '9-pile' system for allocating grades (Ebel 1972:150, also Broughton et al 1978:147). This is a method of applying an instant overall mark to extended responses to test questions, such as

essays or other written answers, either as a once-only spot judgement where an overall impression is all that is required; or as a preliminary to further more detailed marking; or (in a competition, for example) as a means of eliminating answers which are clearly sub-standard before devoting more attention to the rest. The principle behind the system is that it is comparatively easy to decide which of two scripts is more meritworthy than the other, and so to sort them into two groups (the operation known as 'paired comparisons'). In addition, as an extension of this comparative judgement, it should be possible to sort into three groups by deciding in which pile each new case should be placed by comparing it with allocations already made. If these three piles are again sorted into three each, the final result is nine piles on an equalised distribution of merit. The system can be applied in any context of allocating values and is particularly apt for grading difficulty in the present case because it begins with no preconceived notion of how difficulty is to be described. It depends entirely on binary judgements about what is found. Description of the dimensions along which difficulty may lie can be attempted after the allocations, if necessary.

All the entries on the checklists for processing were transferred to small cards (7 x 7 cm) so that they could be used for 9-pile sorting by hand, according to perceived comparative difficulty. Examples are given in *Figure 4.8*.

Figure 4.8: 9-pile cards: 6 examples

sub-heading:	input	sub-heading:	input	sub-heading:	response
content:	rubric alone	content:	travel brochure (tough text)	content:	report to teacher, class
rating:	1	rating:	5	rating:	3
sub-heading:	response	sub-heading:	input	sub-heading:	response
content:	chat with partner	content:	4 texts of 14 lines	content:	homework
rating:	2	rating:	4	rating:	1

Applying this system threw up several rules which simplified the process in this case, for example:

- the entry 'n/a' (= not applicable) was classified as zero, ie pile 0;
- if a pile consisted of two or fewer cards, they were allocated to one or other of the piles on each side, whichever seemed more appropriate;
- the number of piles for any given heading therefore depended on the number of differentiations found to be feasible within it.

It is notable that in this application of the 9-pile system, it was found possible to distinguish only four or fewer levels for 6 of the 23 headings. The most frequent division was into 6 piles (this occurred for each of 8 headings), but in no case could the content under a heading be divided into all 9 levels, as the 9-pile system expects.

The reasons for this may lie in the lack of detail in annotations on which the grading was decided (necessarily short because of space available on the checklist), or, more fundamentally, in the nature of the entities to be graded. For example, the number of 'facts' to be coped with under 'load' produced the widest distribution, with no zero pile because it was relatively easy for the analyser to identify how many facts seemed to be involved in a given task. On the other hand, 'link, with mode' involved decisions about the distance between, for example, spoken and written versions of the same events, a conceptualisation of a particular 'mental gymnastic'.

It is also worth noting that the number of cards (608) amounts to rather less than half the number of entries (1242). The explanation for this is related mainly to the incidence of 'not applicable' entries (n/a), where the event to be recorded either did not occur or was 'given' (rather than to be found by the learner). The total of these zero ratings was 260.

The cards were also used for amalgamation and redistribution under the headings where the results of the practical application of the checklist seemed to warrant them. As a result, the content under the headings was revised, with new keywords covering up to four of the previous subheadings.

The detailed workings of the 9-pile system in this instance, which consists of notes written to record the procedure as it was put into operation, are illustrated in *Figure 4.9.*

Figure 4.9: Notes resulting from 9-piling

These notes were written to record the results from using the 9-pile system to assign values to the concepts represented by the headings on Analysis sheet 2 (AS2). The keywords had by this stage reached their final form, even though the layout of AS2 itself had not. The comments include in the listing:

- the problems arising with the application of each keyword;
- a question which would encapsulate the issue (the learner's 'what am I doing here?') for the analysers' manual;
- an indication of the most demanding ('top') and least demanding ('bottom') examples found in the course books;
- an indication of the grading applied by including the next-down-from-the-top and the next-up from-the-bottom

For each keyword, P = problems, OQ = orienting questions, MP = marker points bottom (▼) and top (▲); ■ = next pile, up from bottom or down from top.

input (35 cards, 6 piles)

P

Complex criteria (x 3). Is R easier than L? Is both together more difficult? Does adding v make it easier or more difficult? More support to meaning search or more to take in?

OQ

How much input is there? & how quick does self have to be in receiving it? & how difficult is it to take in?

MP

▼R phrases, L; L model; all before; LT,v; R headlines
■ R texts; RL(+W); R info, rules; R transcript; Rv nationalities, flags
▲ R phrases,L Sts,T; L opinions; L Sts; L verse (song); R fast(time limit)
■ R texts & gaps; 4v,R2Q; R & m/c; L dialogue & R items; L text; L scripted, half scripted

response (29 cards, 7 piles)

P

Coincidence with last category: this is pressure, that is form/content

OQ

How quick a response is required? or can self take own time to resond?

MP

▼ own time; repeat; when ready
■ copy; model; class chorus
▲ make up mind; guess; summarise; W pairs, S class
■ role play; S ideas; S,W; argue

speak to class (9 cards, 4 piles)

P

18 n/a. Character of self = seriousness of effect - shy>outgoing//stand up>hide

OQ

How far is self exposed to psychological pressure of S in front of peers?

MP

▼ no
■ to group; pairs; in group
▲ yes; choice of T

(continued)...

The final outcome of all this experimentation was therefore a checklist which had been refined through constant trial and revision. It could then be applied with some confidence in the trials of the whole process from course book analysis to assessment system. This final version of the checklist is set out in *Figure 4.10*, and the consequent form of AS2 given in *Figure 4.11*.

Figure 4.10: Processing categories: operational version

HEADING and keyword	subsuming	content/comment
PRESSURE		
input	-	speed, amount, kind
response	-	speed demanded
	speak to class	exposure
	competition	game
load	facts	how wide ranging
	participants	how many characters, students
	together	synchronic
COGNITIVE PROCESSING		
search	invent	dig out of imagination
	logic	dig out of logic
	remember	dig out of memory
link	with mode	difficult to realise in coding, to differentiate
	with info	
	to ideas	
	to situations	
COGNITIVE FAMILIARITY		
know	connect	schemata and fillings
	general knowledge	
	already	
	accessible	
individual	self	important for individual
	emotion	
OUTPUT		
demand	open/closed	2 aspects: form & content
	system	
	formal/informal	

Figure 4.11: Analysis Sheet 2

Analysis Sheet 2: processing

11	12	13	14	15	16	17	18	19
task (as col 1)	input	response	load	search	link	know	individual	demand
	16	*15*	*17*	*16*	*16*	*16*	*13*	*16*

An evidently viable system had now been set up for both AS1 and AS2. The next step was for a set of about 40 analysis sheets to be completed without variation by the researcher and a further 10 by colleagues - staff and postgraduate students. These more constant applications demonstrated that a reasonable level of agreement was feasible, but showed that further guidance was needed for analysers, and a more comprehensive manual was therefore written to explain in greater detail how the analysis was to be carried out.

By this time the material needed for any actual analysis to be undertaken had reached such proportions that it had become difficult to handle, partly because of sheer bulk. The analyser needed to refer to the student's book, the teacher's book, lists of what it was permitted to fill in on the analysis sheets (three pages at that time, since analysis sheet 2 contained examples which extended it to a second page), examples of past applications, supplies of blank forms, and now a manual of eight pages of closely-packed information, including explanations, examples and a glossary of terms which defined every word used in the system. All this paperwork needed constant cross-reference before each decision could be arrived at, a turning of pages to and fro which increased the time and temper expended on the exercise.

4.5 Computerisation of analysis system

Some neater access was needed to replace such an unacceptably cumbersome procedure. This was provided by introducing computer processing. HyperCard (© Apple Computer Inc 1988) is a computer-controlled filing system which represents multiple stacks of cards which can be instantly accessed as required. This application of the HyperCard program was designed as a sequential procedure which led the user through the process of completing AS1 and AS2 by making available to him three kinds of information: administrative instructions, guidance on what to enter on the Analysis sheets and a glossary of the terms used. Three different kinds of card were therefore designed, differentiated by layout and by the use of italics and varying fonts, so that it was clear what kind of information was being consulted at any given moment. The total stack amounts to 118 cards.

The procedure is self-explanatory, in that once the program has been initiated, the user needs only to follow the information given on successive cards in order to complete the Analysis. In addition to the program and a computer to run it, he needs a sample of a completed Analysis sheet, blank Analysis sheets for completion, the course book (both student's and teacher's, if existing), and a pencil.

Every card includes 'buttons', icons which can be activated with the mouse. Two of them are labelled respectively with forward and reverse arrows (⇦ ⇨), and these effect a move to the next card in either direction. The other icons, in the form of a circle or a square with words attached, initiate longer jumps: to the next or last heading, or to the exit from the program. Within some cards there are words or phrases highlighted in bold, and these also act as buttons to access definitions from the glossary.

First, some examples of administrative cards. *Figure 4.12* is the explanation of the system as it appears near the beginning of the procedure, with chapter headings ('heads') allowing the analyser to find different information as required (*Figure 4.13*). The heads relate to different aspects of the scheme, some to the working of the card system itself, some to the entries to be made; and finally there is the glossary of all terms used, which avoids multiple page-turning by making all definitions available at the command of a key.

Figure 4.12: HyperCard: card system

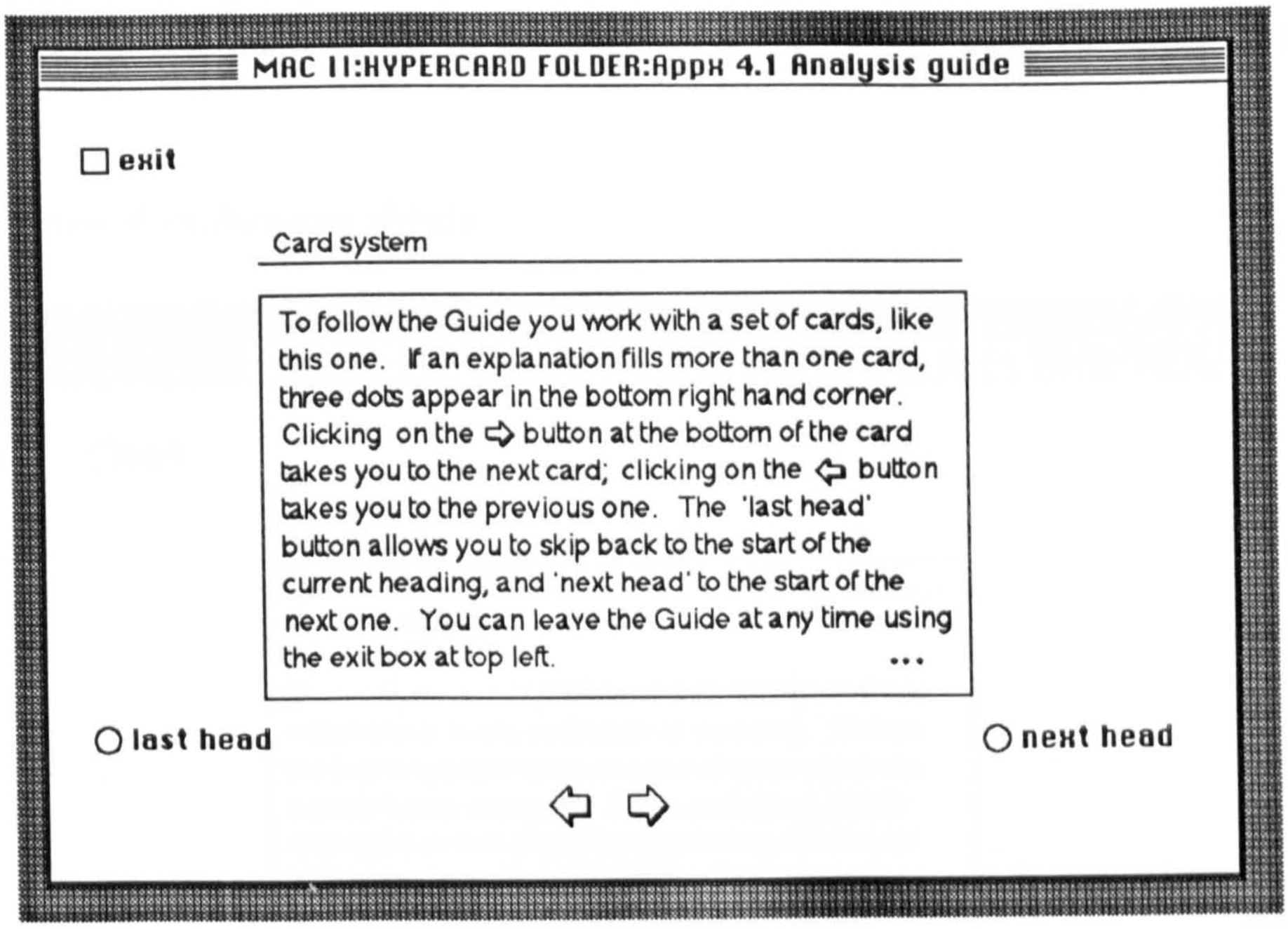
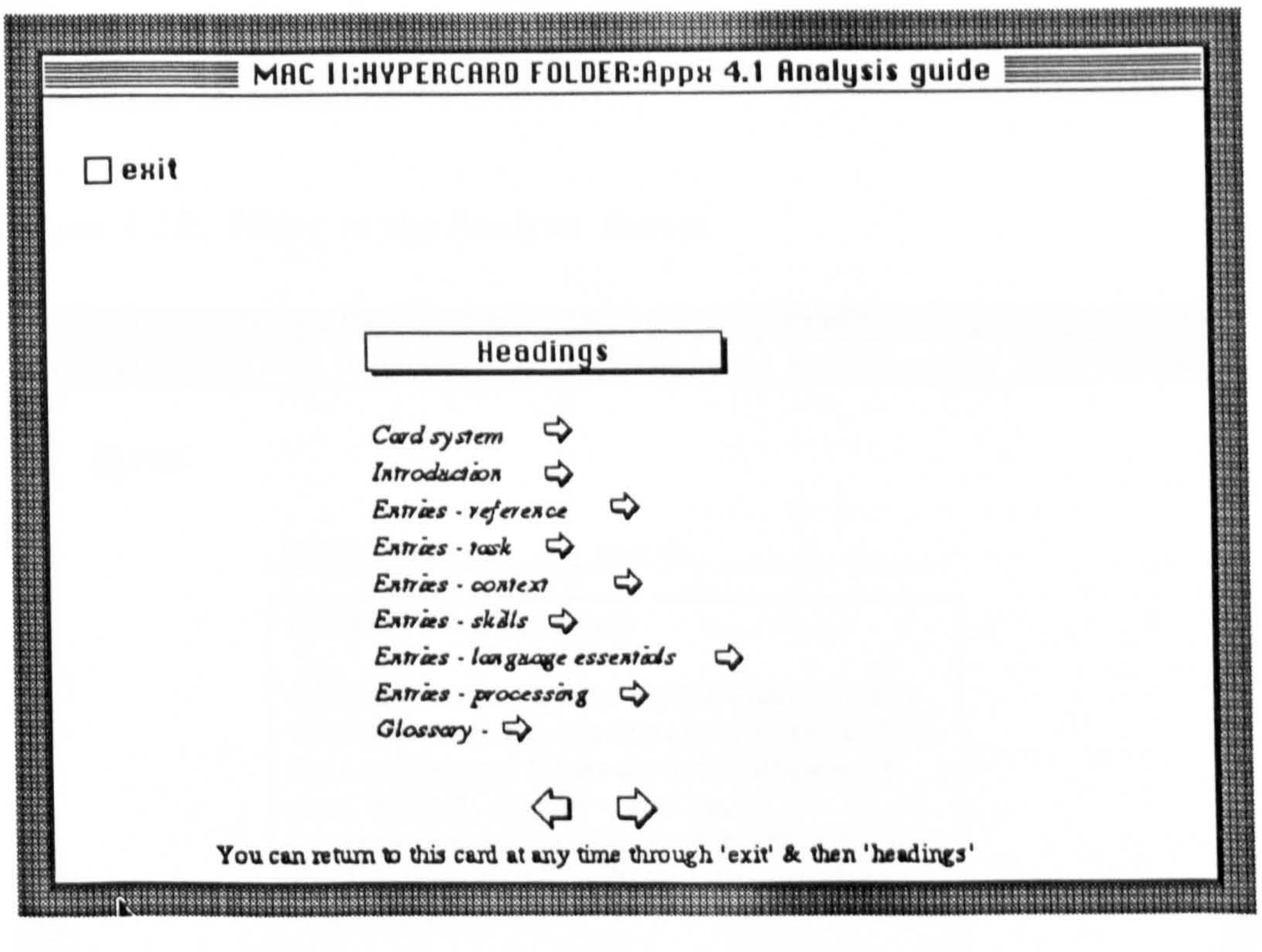


Figure 4.13: HyperCard: headings



To begin with, the analyser follows the sequence of cards straight through, referring to the glossary when necessary by means of the active 'buttons' which instantly bring up the necessary information and allow instant return to the point at which the sequence was left. But in due course (in practice, quite quickly) he can skip information he knows and search for information only when he needs a reminder or when there is too much to remember. The scheme is further simplified by the

inclusion of all examples for AS2 in the program, so that AS2 itself reverts to one page only, making it simpler for analysers to use. The following six figures are examples of cards which demonstrate the working of the system for AS1.

Figure 4.14: Analysis sheets

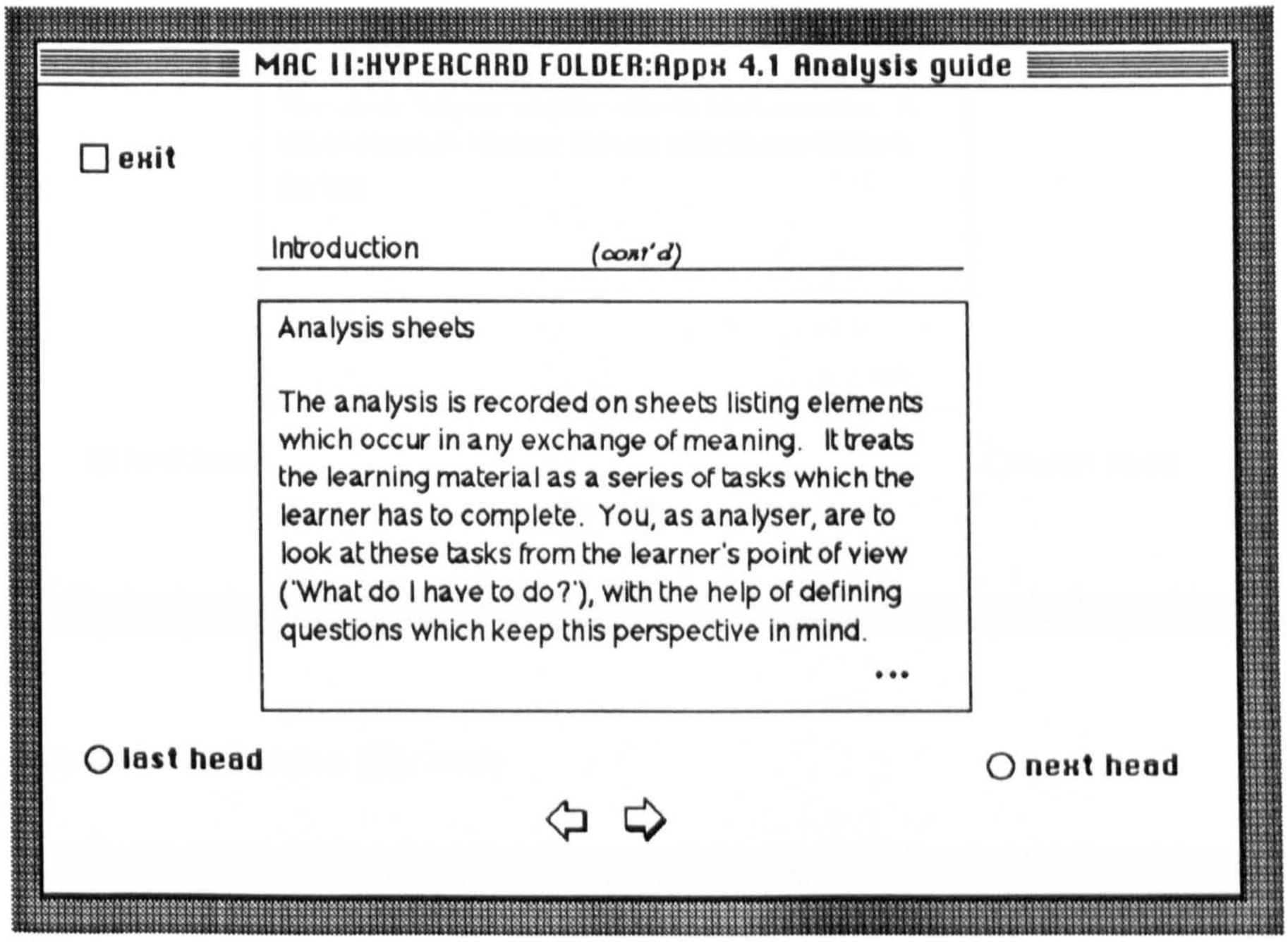


Figure 4.15: Filling in the Analysis sheets

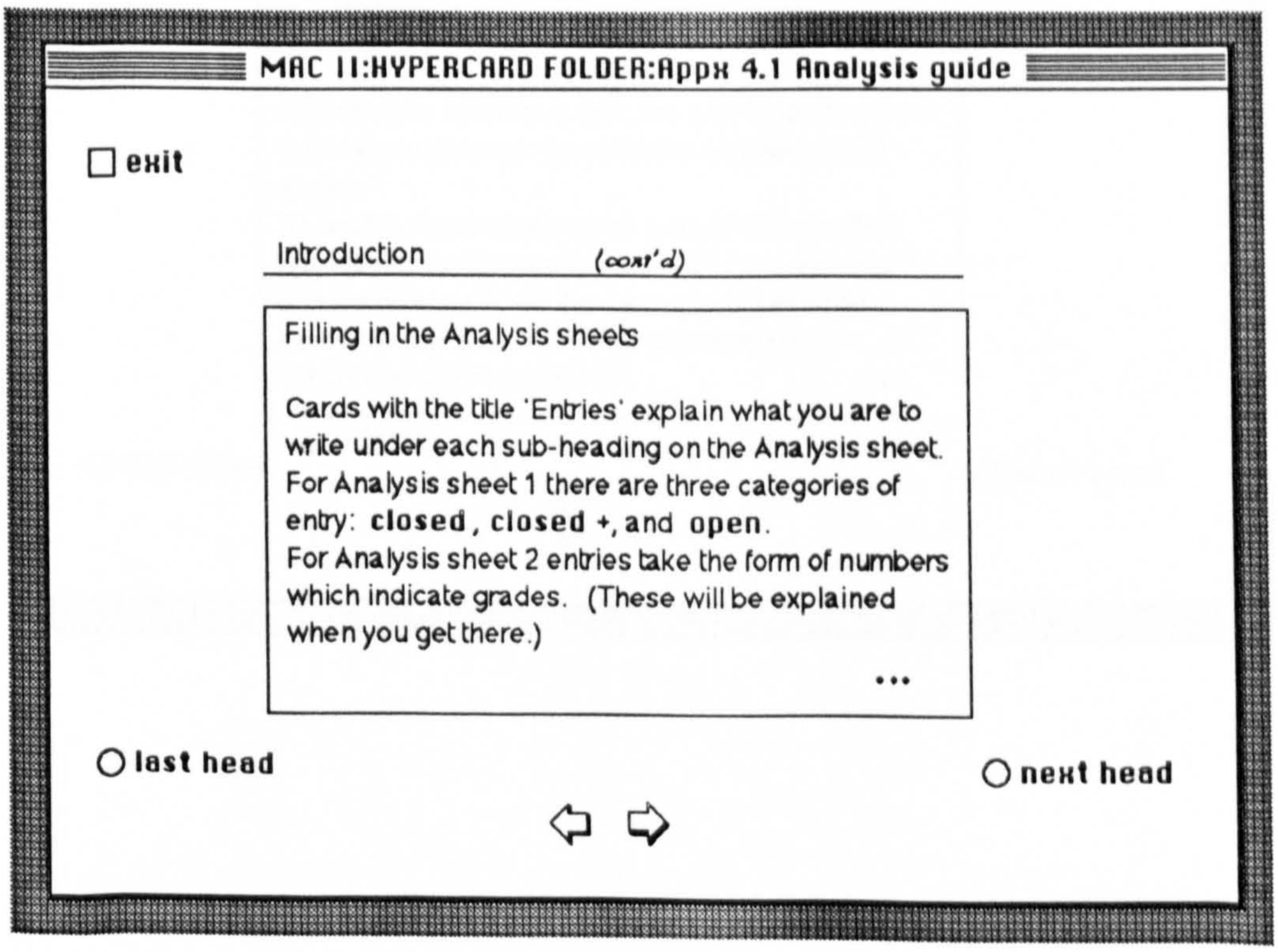


Figure 4.16: What to write

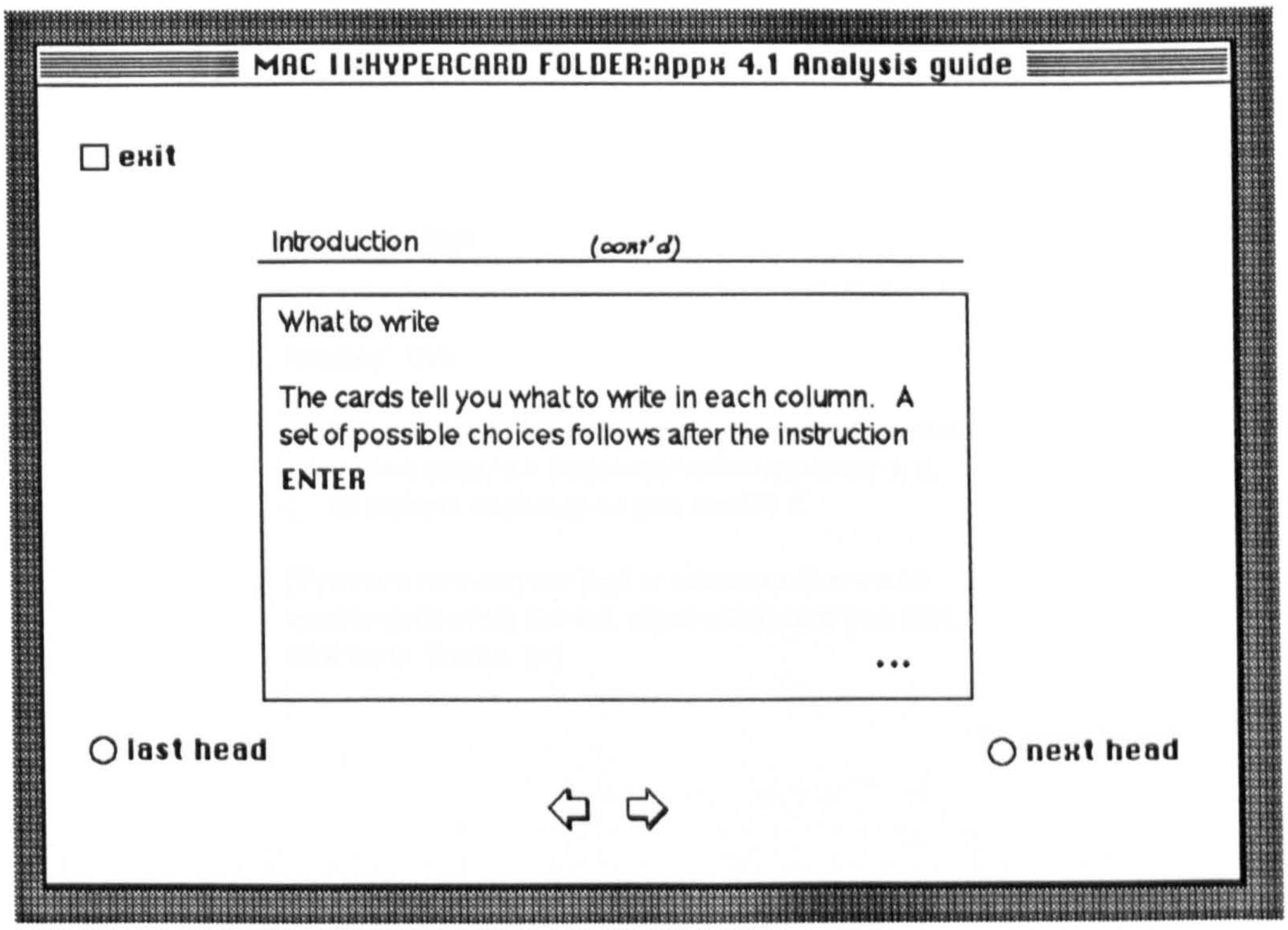


Figure 4.17: Entries (2): task

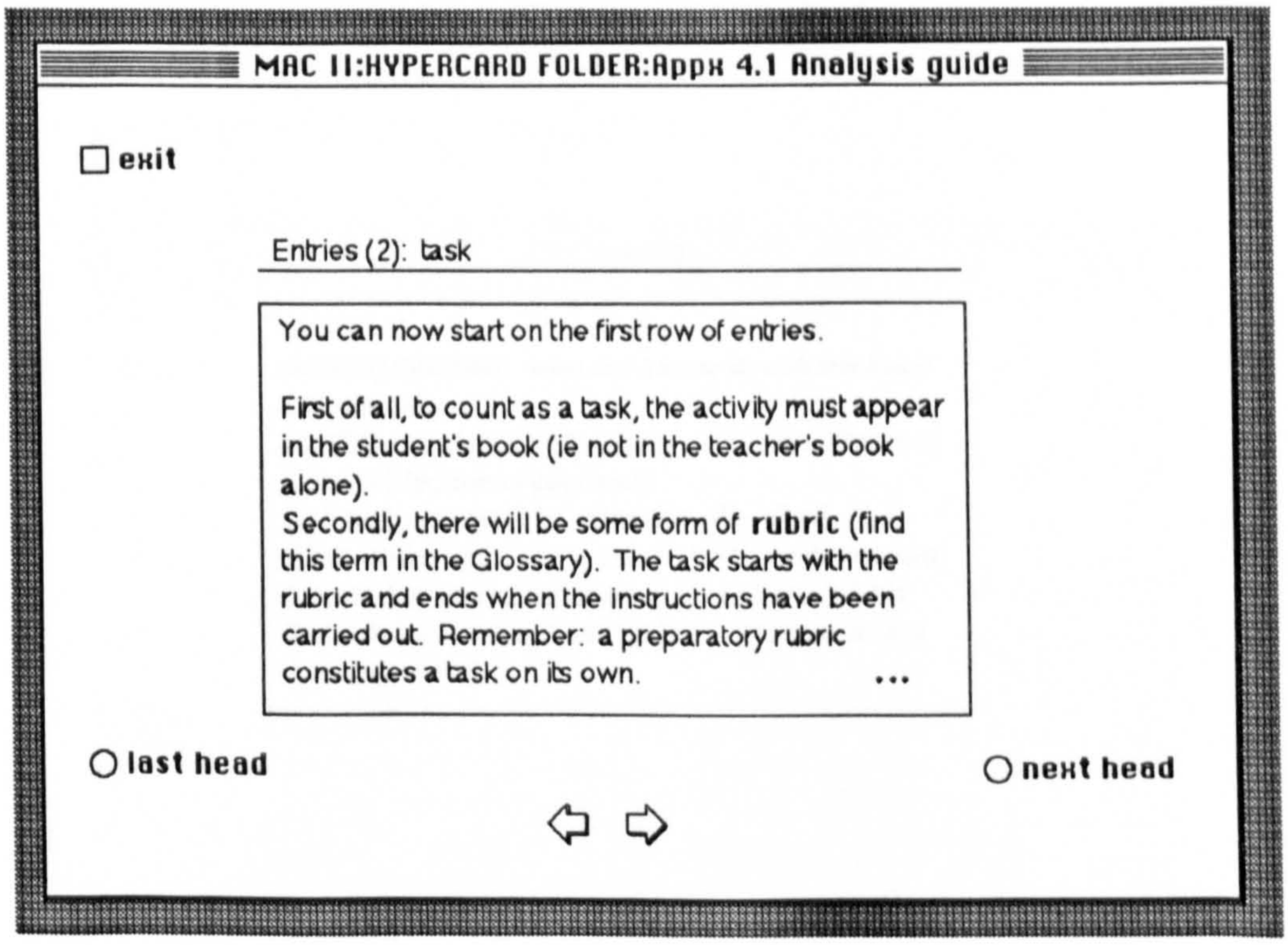


Figure 4.18: Entries (3): task (cont'd)

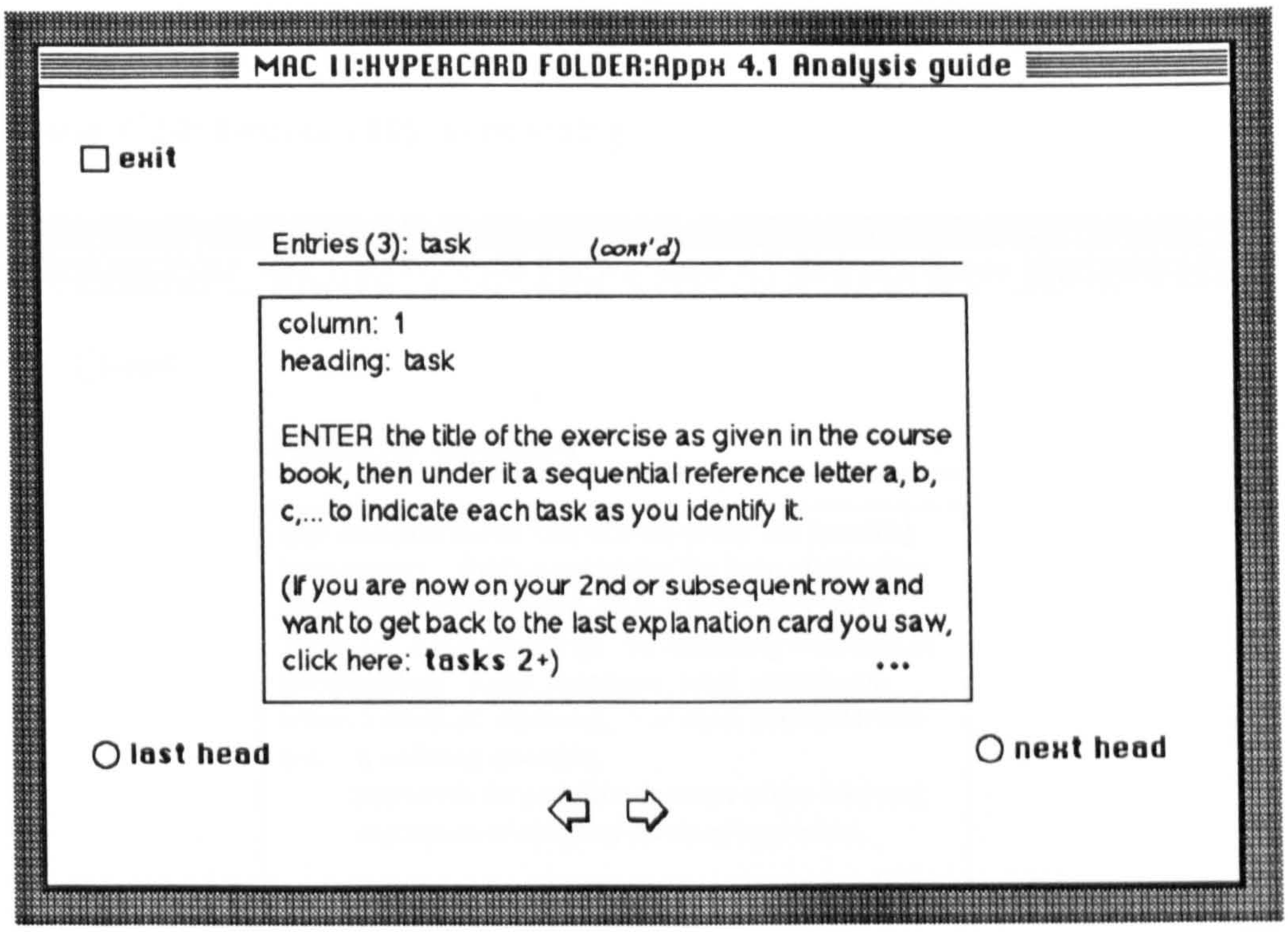
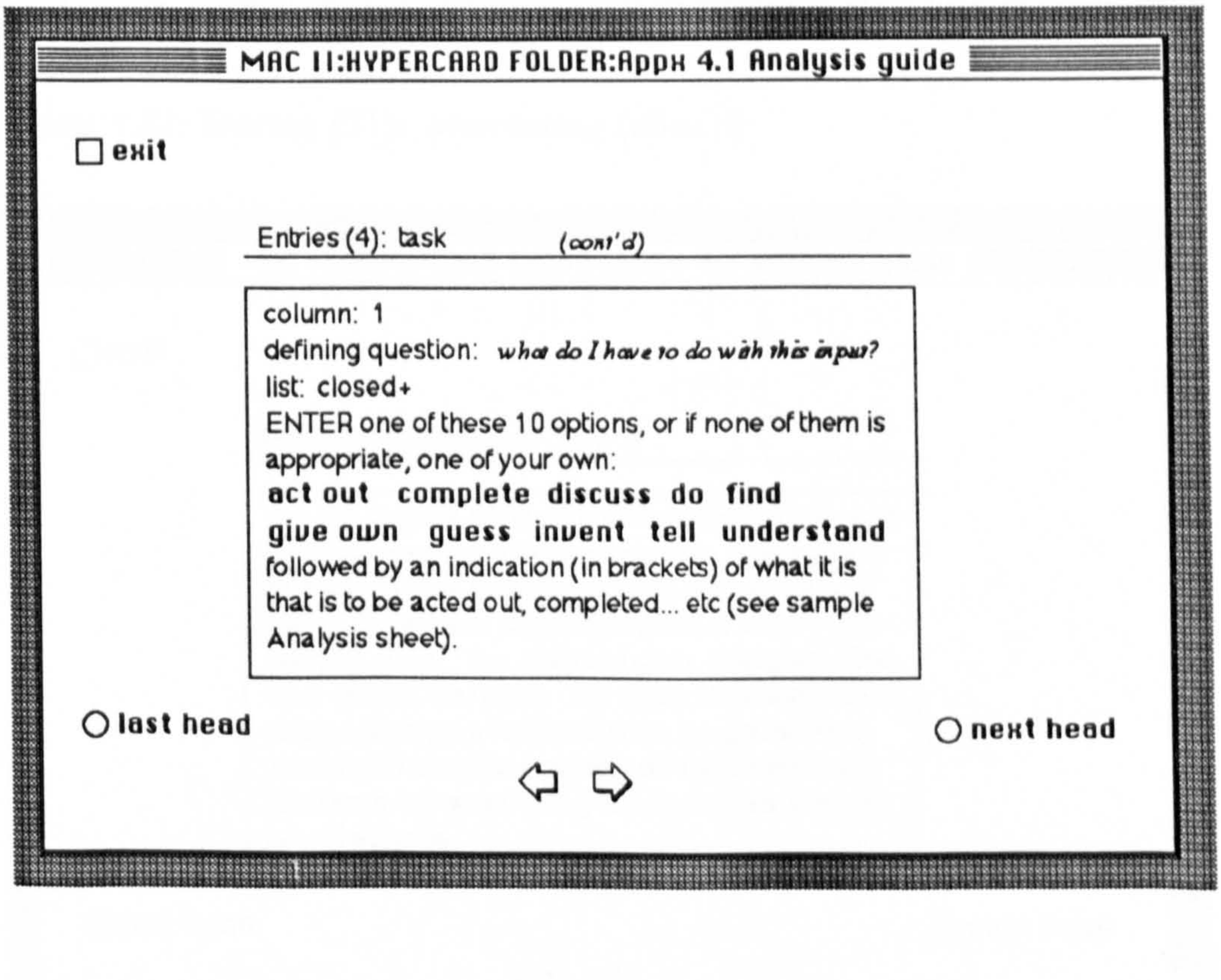


Figure 4.19: Entries (4): task (cont'd)



The system for analysing processing on AS2, as explained earlier, is different from that for AS1, and this is shown by Figures 4.20 to 4.23.

Figure 4.20: Entries (20): processing

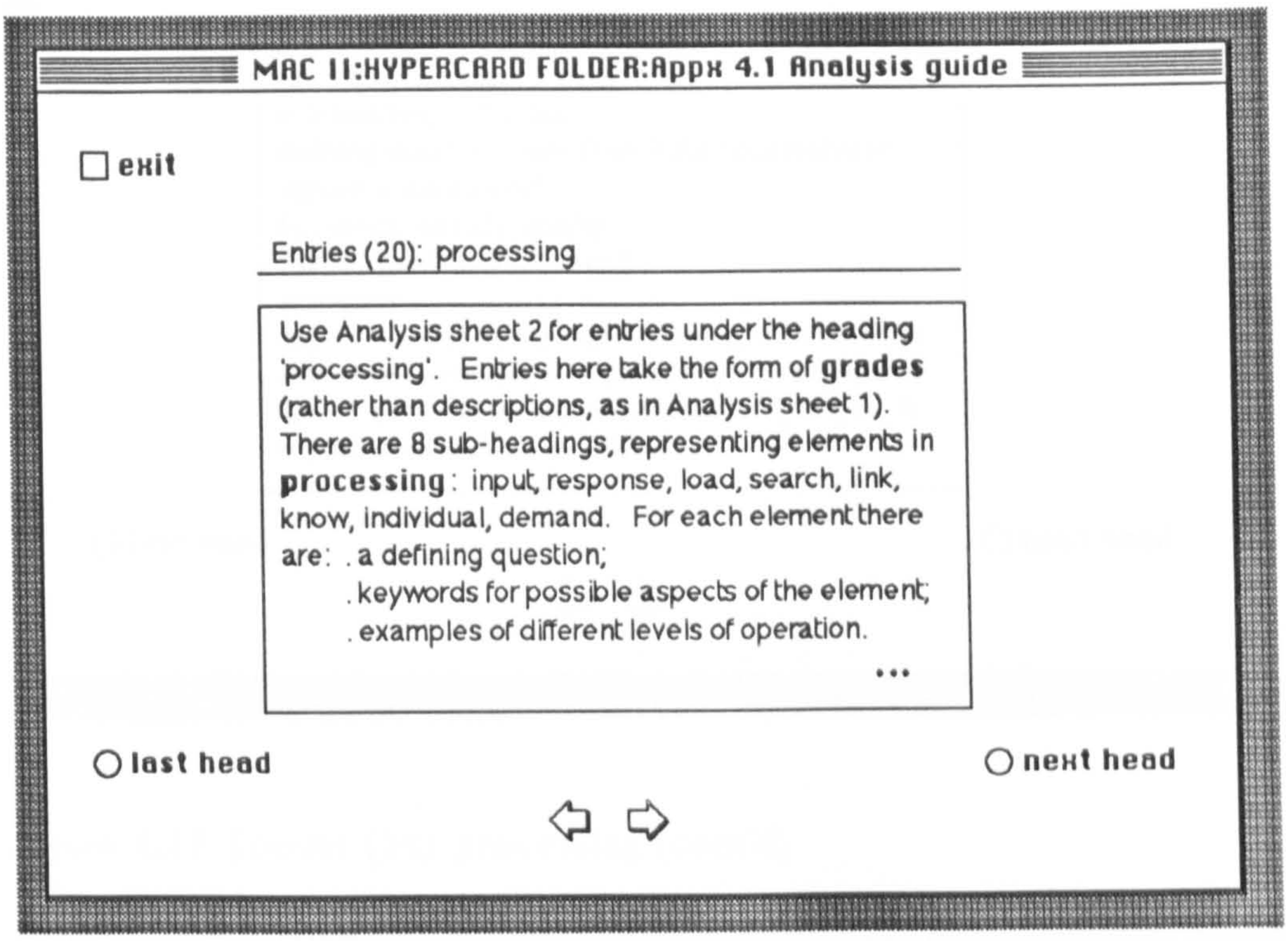


Figure 4.21: Entries (21): processing (cont'd)

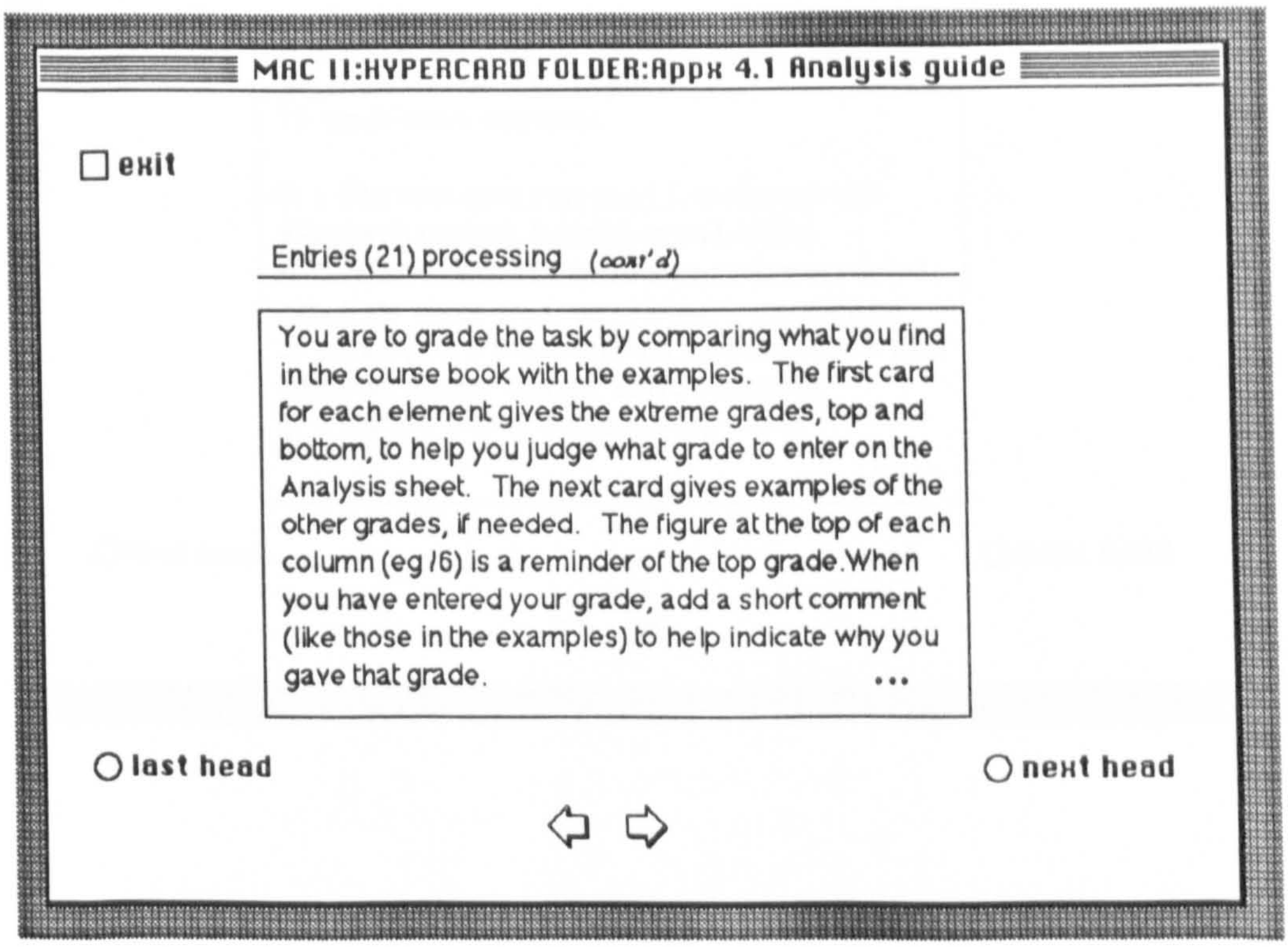


Figure 4.22: Entries (23): processing (cont'd)

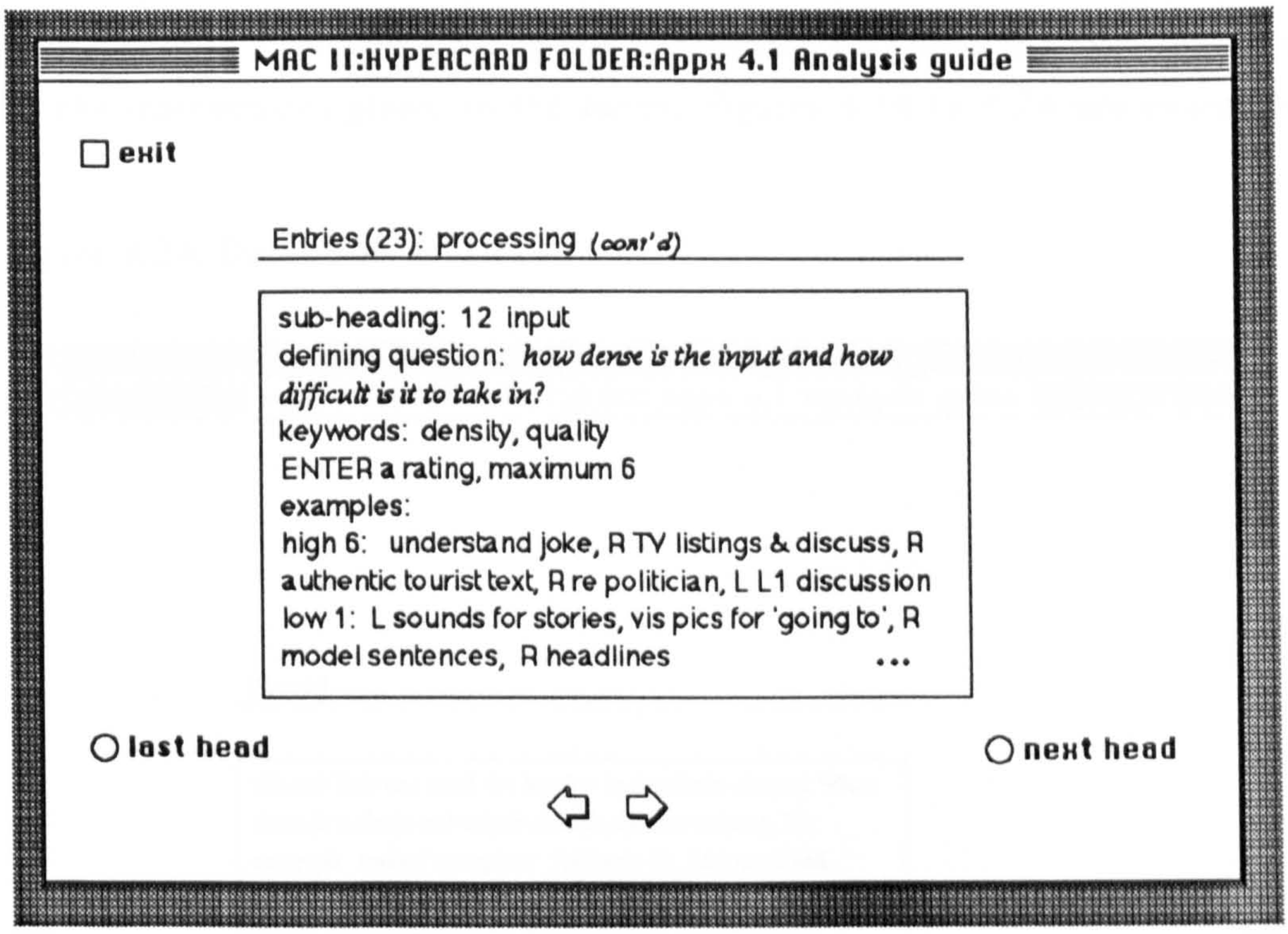
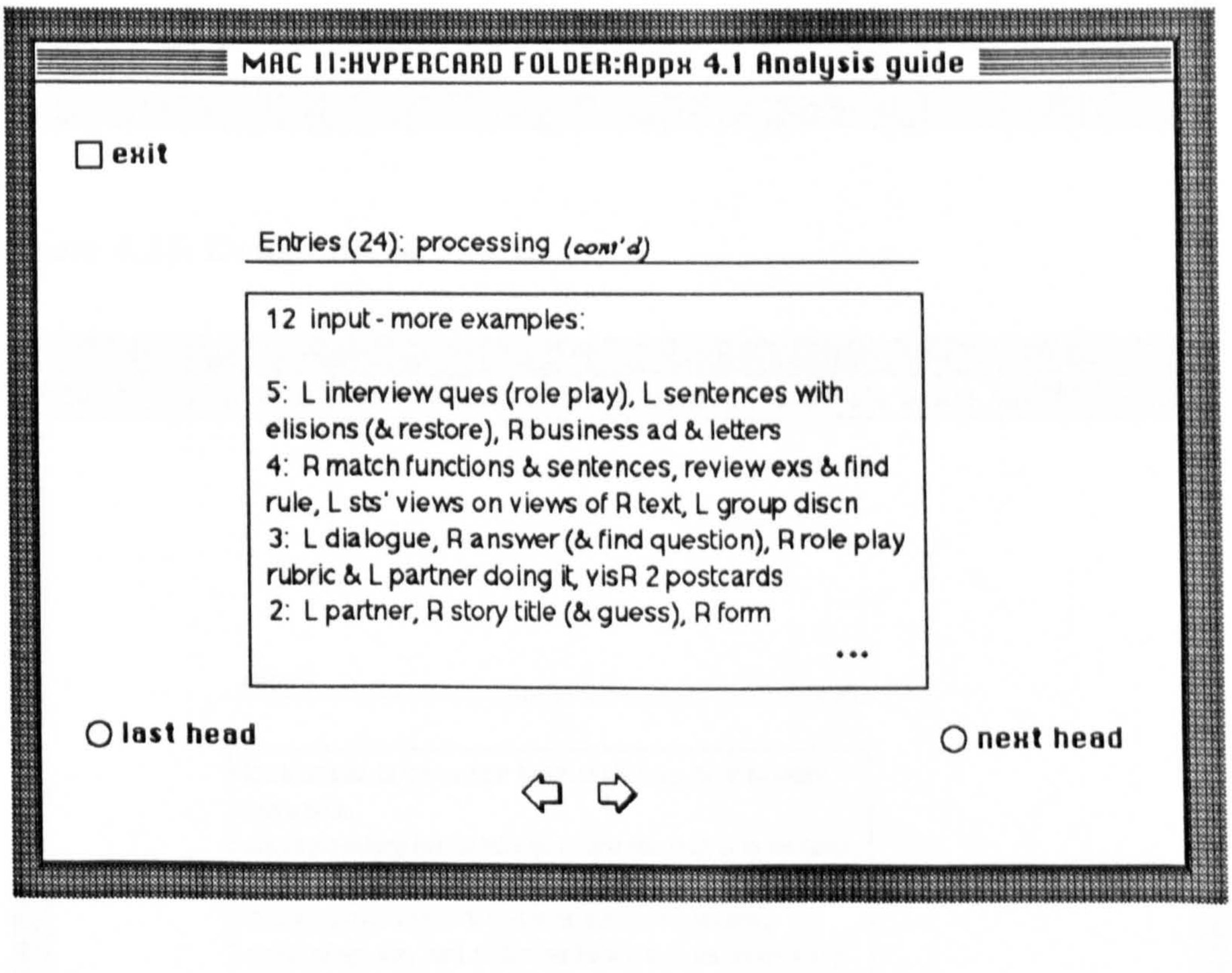


Figure 4.23: Entries (24): processing (cont'd)



All these cards headed 'Entries' belong to the set which gives guidance on what is to be entered on the Analysis sheets. A third kind of card is used for the Glossary, which defines all words used in the application and links them by means of 'buttons' to the instructions given on the cards. *Figures 4.24* to *4.26* are examples of these.

Figure 4.24: Definition of 'closed'

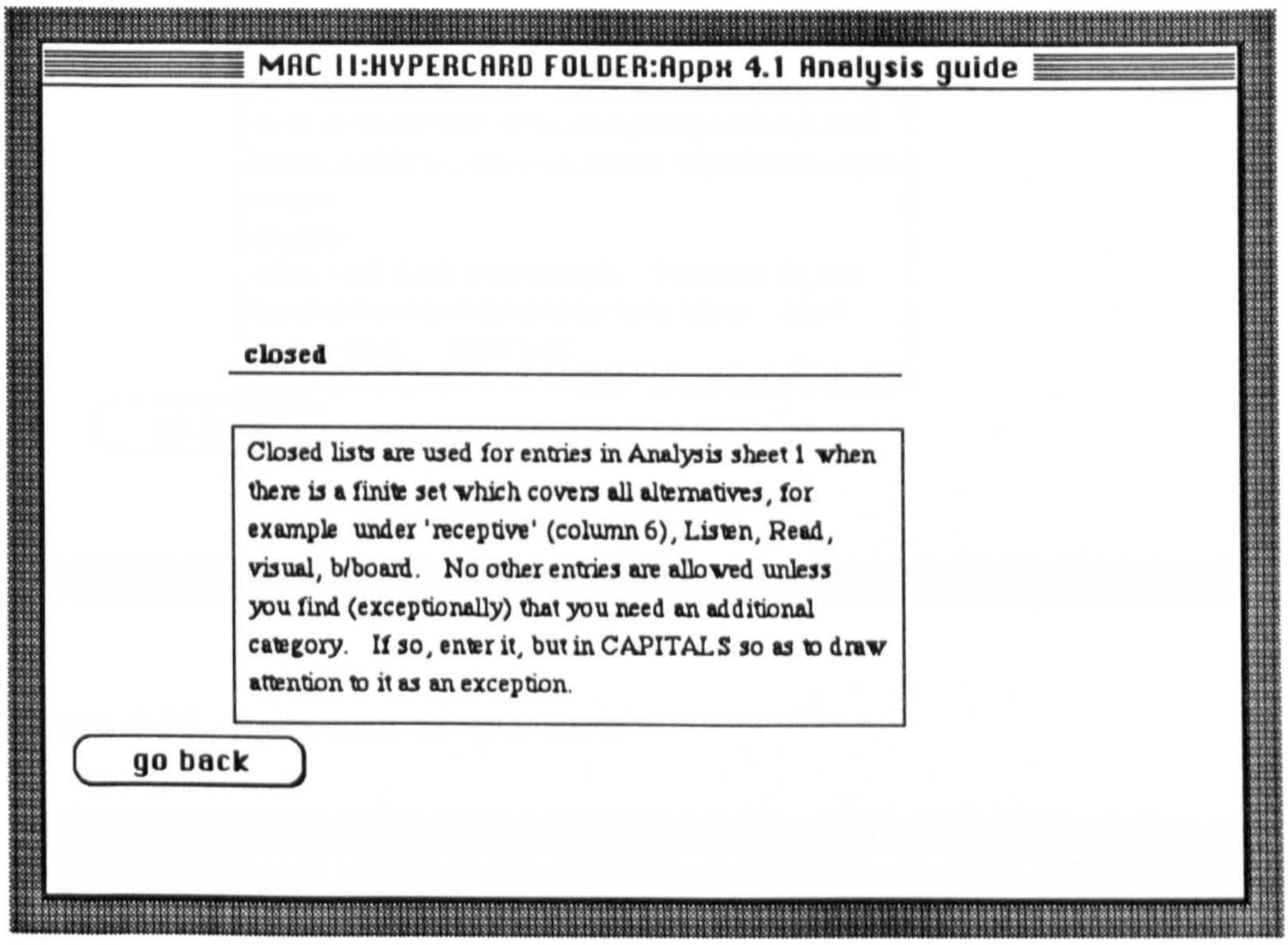


Figure 4.25: Definition of 'rubric'

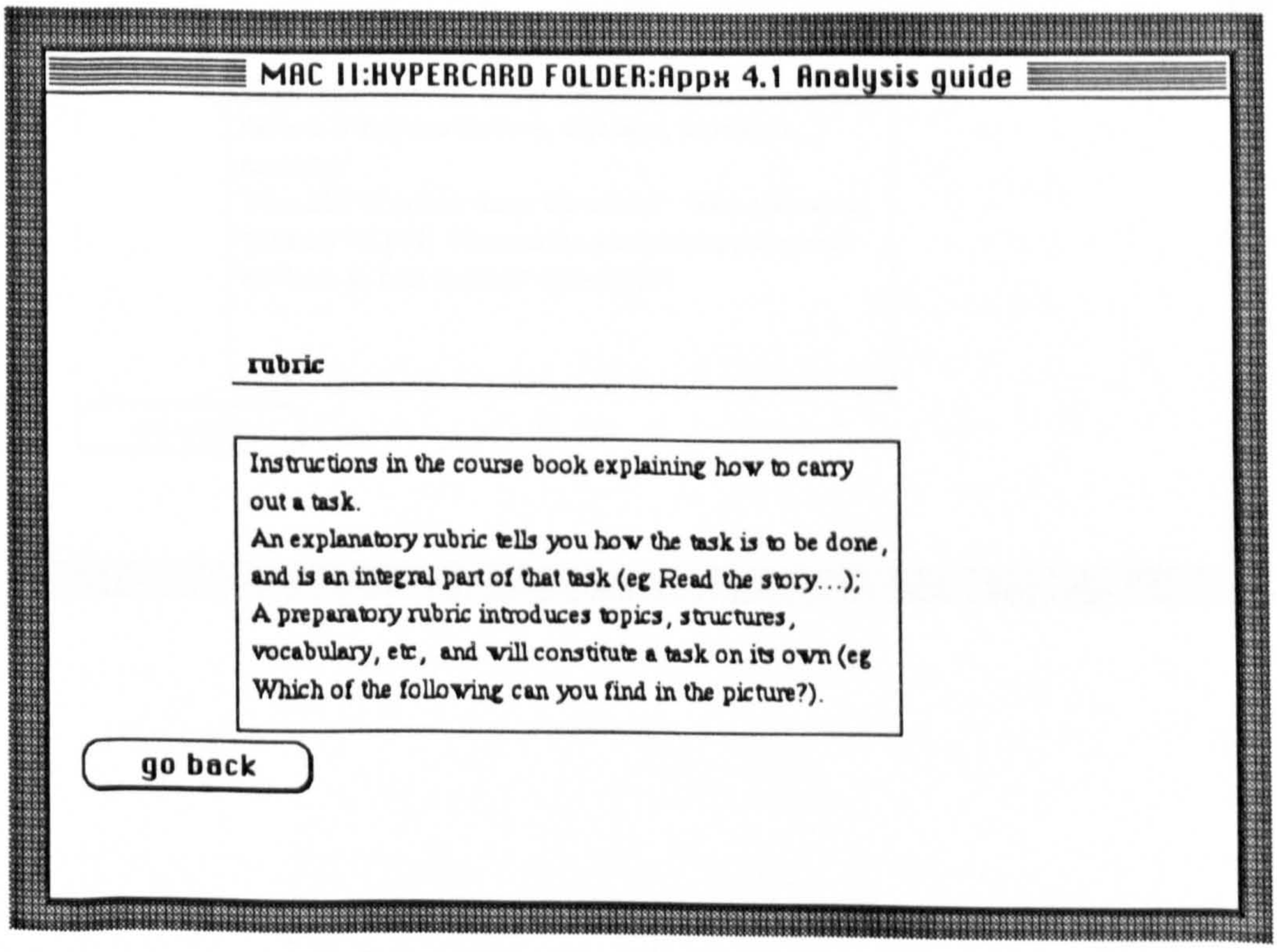


Figure 4.26: Definition of 'author'

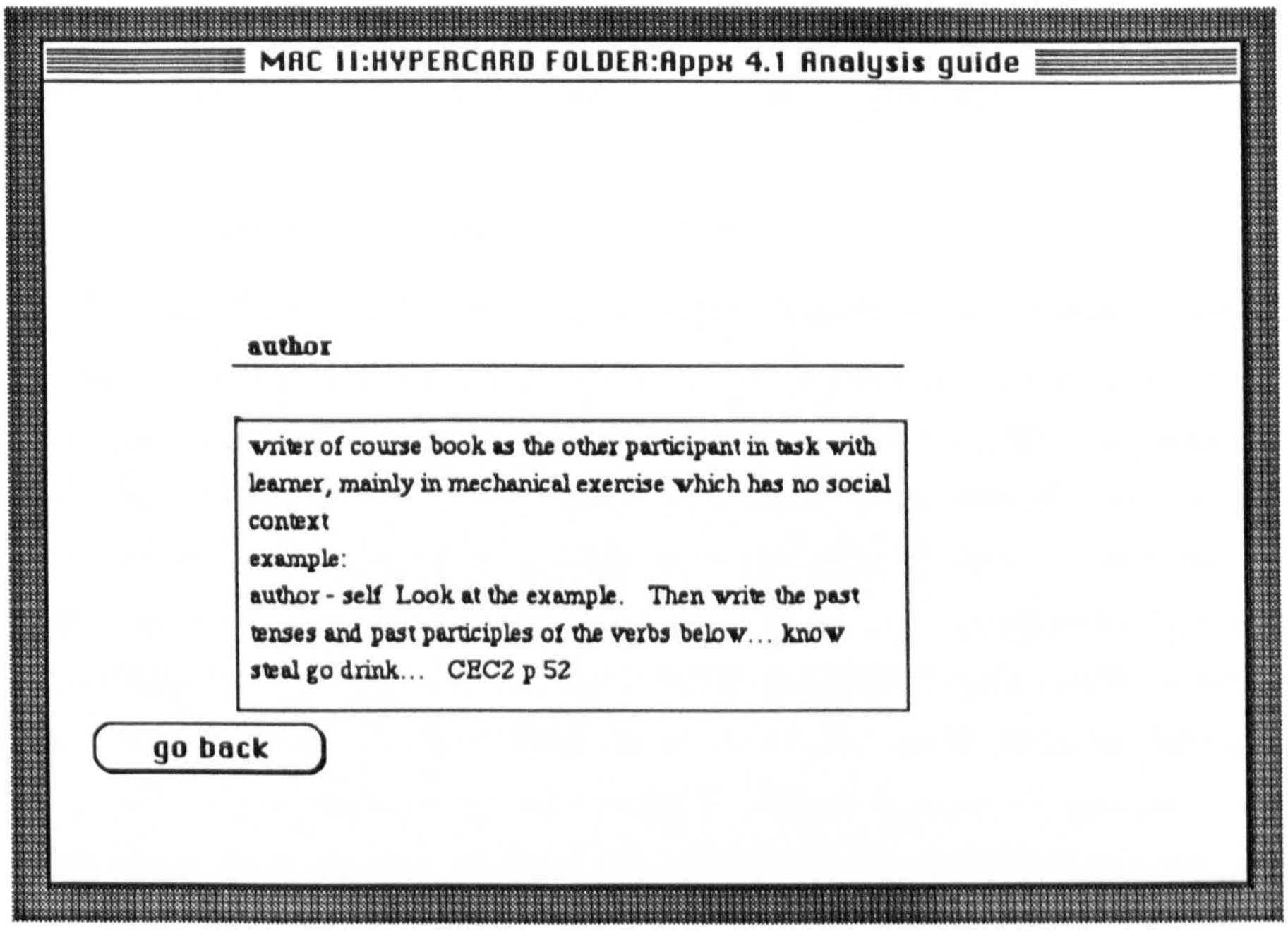
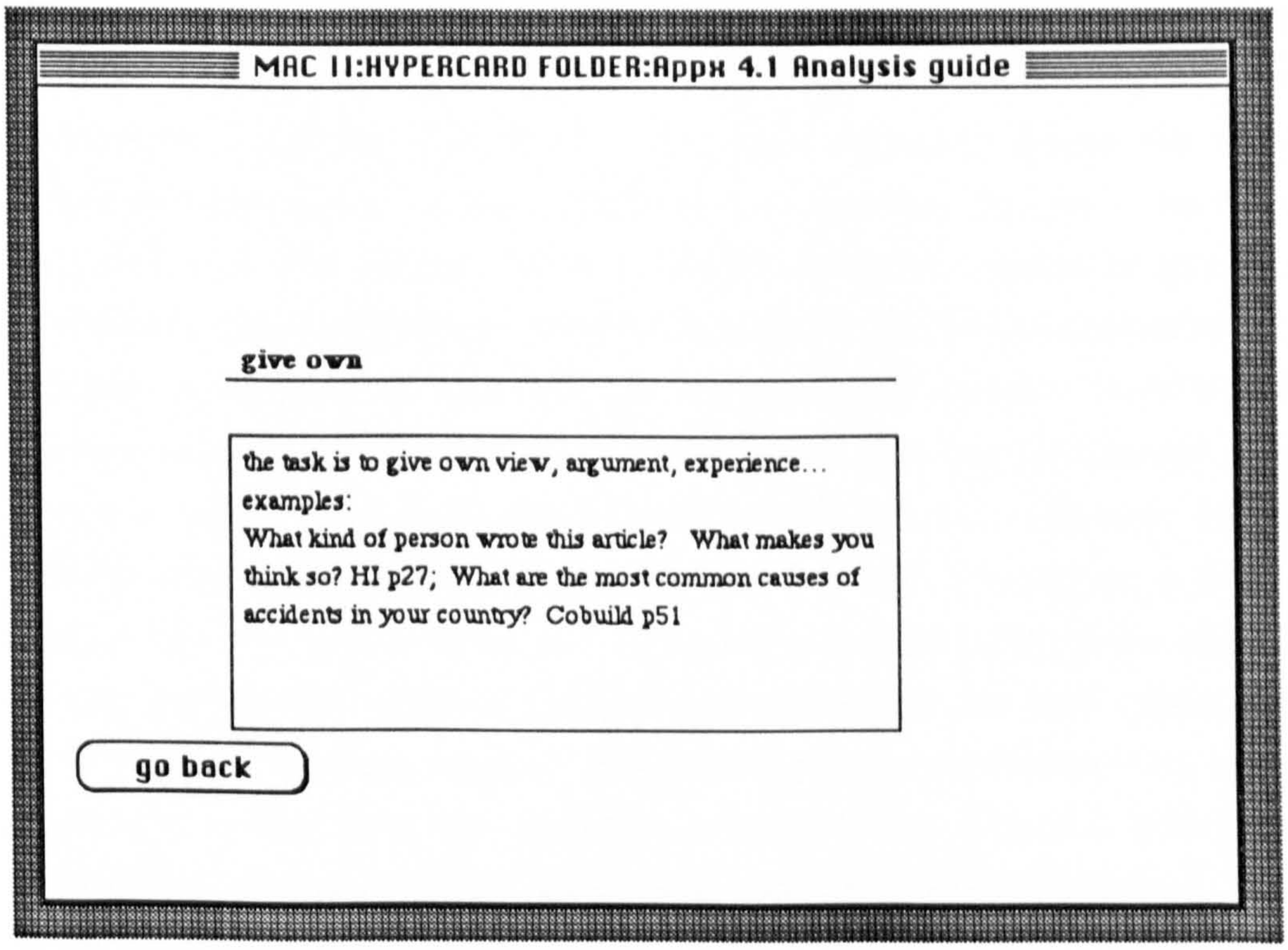


Figure 4.27: Definition of 'give own'



These extracts from the HyperCard stack illustrate the way in which the system operates to lead the analyser through the analysing procedure with information and guidance on how to complete the Analysis sheets. The full printout in Appendix 4.1 indicates the extent of the coverage and detail provided.

4.6 Responses to the system

At this point, it may be useful to attempt a summary of findings from discussions of the AS1 and AS2 in final draft forms. There were two main trials with colleagues, one before and one after computerisation. On the first occasion, participants worked with the same examples from a course book, and discussion covered several areas of doubt which were thought to need clarification. Detailed revisions were suggested to make the manual easier to use and the intentions clearer, for example rewording some of the defining questions, emphasising the difference between 'load' and 'input', adding a noun after 'demand' in the 'task' column, filling out the code where it was unnecessarily abbreviated. More important was the common feeling that there was too much detail to cope with, but there were no suggestions for deletions - all the areas included seemed to be necessary for adequate coverage.

But the main criticism was encapsulated in the question: 'who is this learner?'. It was argued that the wide contrasts between student situations in various parts of the world, and the considerable differences obvious even between learners in the same classroom made a 'What am I doing here?' stance too individual for it to be adopted as a working hypothesis. Since this comment denies the whole principle of the analysis, some counter justification is needed. For AS1, the facts sheet, the approach was intended to define a task by means of a series of questions referring to the activities required to complete it, and the learner, no matter who, faced these demands as required of a learner, not this particular learner. The 'I' who has the task to do is confronted by a situation; the task certainly has an affective impact on 'me', who may react to it from a cultural background or with a personal emotion (eg like/ indifference/ dislike) which colours 'my' attitude, but this makes no difference to the facts of the task and its demands. The previous experience of the particular learner is not at this point relevant to the requirements of the task: the analysis records what is required of any learner. For AS2, the judgements sheet, the same arguments apply, but in this case the question is about what makes a task more (or less) difficult for a learner rather than about what it consists of. Again, for example, it is obvious that some learners enjoy standing up in class talking whereas others cringe if asked to speak on their own, but these different personalities would agree that haranguing the class is of a different order from talking quietly to a familiar partner (processing 13 - response). But this criticism of the approach from 'I' was not universally made: most of the contributors found no difficulty in applying the principle throughout, and some found it positively revealing.

At the second trial, which was concerned with the working of HyperCard, there was near universal agreement that the computerisation of the system made it simple to carry out, though there were still a few reservations about detailed wording. In the main, the system as such was considered to work well and to be a feasible means of course book analysis, if still highly detailed and quite complex in operation.

The extent of the work with assessment sheets is indicated by the following account of development. Several variations on a tabular format were developed until the first trial version was arrived at. 40 examples of this version were completed (approximately 3500 entries) before the revision was made to separate out 'processing' from the other categories. This was the point at which AS2 was introduced. In the first version it was accommodated on two pages because examples were included to help analysers decide what to enter in the appropriate places. This was the version used by discussants at the review, but was considered too clumsy. All examples were subsequently incorporated in the manual, which immediately became a data base, for at this point the computerised version of the system was introduced and more than 30 analysis sheets (now double sided, with AS1 and AS2 back to back) were completed as the consolidating run (without developmental amendments to the system), giving rise to about 4500 entries. The trials with schools (Chapter 6) amounted to a further 30 sheets (4000 entries). The 9-piling exercise was carried out twice, once to set it up, using 600 cards to work out how the system would work for the allocation of difficulty levels to categories of processing, then again in a consolidating run of 900 cards to confirm its viability and to provide examples for the manual.

All this accumulation of instances was intended to establish the consistency of the system, firstly by discussion (an interaction with the materials, in effect) to guide decisions on what was to be found; and secondly, when the operation of various parts had been consolidated, to demonstrate that no further changes need be made to the system for it be applied consistently to whatever learning tasks it might be asked to analyse. As Anastasi points out (1968:84), homogeneity is an important factor in the reliability of a test: in the same way, the extensive use of the analysis system provided cumulative evidence that it could be consistently applied. This repetitive procedure also helped to ensure that anyone competent to do so could operate the analysis according to the manual. The next question then is: how far does the system work consistently among analysers? Of the two kinds of rater consistency (intra-rater, that is each analyser within himself, and inter-rater, that is similarity between one rater and another) the more important aspect in this instance is the former. The equation of one analyser with another is less vital, because the ultimate goal is for a teacher to operate the scheme for his own classes, or at most within the same school, and the particular nature of the situation of that class or school will inevitably affect the way in which the scheme is used. The most important conditions for the consistent use of the system, as with any marking scheme, are the clarity of the rules by which the scheme operates and the

accuracy with which analysers apply them. What the decisions mean in abstract or 'philosophical' terms is in the end less important (since possible interpretations are theoretically limitless) than agreement on their use.

The other important aspect of reliability for the system is that its operation should be generalisable. The analysis system as it now stands is capable of being applied to course books of different kinds at various levels, and of providing useful input for the preparation of assessment specifications which are thus defined by the content of the course book. Its strength therefore lies as much in its content as its consistency. The successful identification of what is required from a learner by a course book and allocation of the results into identifiable categories means that the system can be applied to learner activities of any interactional kind, and this, if interpreted in the ways which have been discussed in this chapter, means universally. At the same time, it has the potential to relate to the individual interests and the needs of each learner, provided a way can be found of relating the work done in class to some productive procedure which will give him the opportunity to demonstrate what he knows.

The system is also administratively efficient: the use of HyperCard brought the application of the scheme down to manageable proportions and there seems no obstacle to its practical use by any qualified and interested party given the will and the equipment. It is tempting to suggest, though the idea is probably unacceptable for the traditional demands of reliability, that what actually appears on the analysis sheets is in the end less crucial for the successful working of the scheme than the use made of it.

4.7 Summary

The search for an analysis system which could represent the content and demands of a course book was made by taking the learner's point of view in setting up categories to record the factual content of the materials offered. A separate set of categories was then devised to record elements of the processing required for the tasks and to estimate their difficulty levels. The procedure followed was to try out various forms of categorisation on a variety of course book materials on an iterative basis, refining the analysis to meet problems as they arose, and then finally undertaking a run without amendment to consolidate the system. At a given point it was decided that the paperwork involved was too cumbersome, and a computerised version of the system was devised to reduce the load on the analyser. At several points in the procedure the system was tried out with colleagues, who reported difficulties, none insurmountable, which were then worked through.

The next step in the project is to find some way of converting the findings of the

analysis system into a relevant specification for an assessment which will provide the basis for the necessary interaction. This is the task for Chapter 5. The following step is to provide a means of realising this interaction in practice as an event in which learners can come together to work towards a fulfilling conclusion, using the language they have available for purposes which are to them self-evidently useful. This is the task for Chapter 6.

Chapter 5

Arriving at a specification

5.0 Introduction

The analysis system is one link in the chain connecting learning with assessment. After the account of its development given in the last chapter, the next links need to be described: how the system is applied to a particular course book, and how this analysis may be used for a specification which will lay down guidelines for the assessment. This chapter therefore has two centres of interest: the operation of the analysis system in practice and the use of the resulting data to produce a specification. It begins with a demonstration of the way in which the materials available to the analyser - the student's book, the teacher's book and the analysis guide - are consulted to arrive at a description of the demands of the course book as entries on the Analysis sheets.

The second section discusses ways of arriving at an assessment specification, and then condenses the data resulting from the current analysis for this purpose. But this is not all: it was argued in earlier chapters that recognising the learner as an individual and incorporating various forms of authenticity were important aspects of the proposed assessment system, and this suggests that assessment tasks should relate to future situations which are plausible for the learners in their physical and social context. To meet this requirement, which stands beyond the learning situation proposed by the course book and its analysis but is nevertheless an important element in the specification, the circumstances in which the learners are currently placed is to be taken into account and added to the summary of learning provided by the analysis.

5.1 The analysis in operation

For the following demonstration of the analysis as it was applied in practice, extracts from the consolidating run mentioned in Chapter 4 will be used as examples. The three course books for this extended application of the analysis system were: *Cobuild* 2 (Willis & Willis 1989), the *Cambridge English Course* 2 (Swan & Walter 1985) (*CEC*) and *Headway Intermediate* (Soars & Soars 1986) (*HI*). The first of these, as reported earlier, was designed as a communicative course based on a corpus of 'real' English, and the second two were highly successful and hence widely-used courses which evidently represented for many teachers and schools the best available.

Sampling

It was clear that some kind of sampling of content was required, since applying the full analysis to entire course books would be too detailed for research purposes and even less acceptable for applications in the field, because too time-consuming for practical purposes. As in all sampling, the main consideration was how far the sample to be decided on was representative of the whole (Lewis 1972:97), which in this case involved decisions on the frequency of Analysis sheets in relation to the course book as a whole and what range of content was to be included.

The first decision on sampling was to analyse the first third of each course book, since those which did include testing material tended to offer three sets of it at equal intervals. Analysing the first third therefore offered the possibility of comparisons with existing testing and parallel applications in due time. Course books are usually set out in units, each of which contains a series of exercises, often following on over several pages, but usually starting consistently on either a left-hand or a right-hand page. In any case, authors tend to follow a fairly standard pattern within a book, probably to reduce the need for constant explanation and also to encourage familiarity and hence security for both teachers and learners. For example, *Cobuild* tends to start with listening or reading with some kind of comprehension exercise (eg note-taking), followed by discussion in class, then a more open-ended exercise for pairs or groups. So as a second principle, it was decided that in order to reduce the effect of patterns of this kind on the sampling and at the same time to enhance content validity, the analysis should be applied every 5 pages, thus starting on left-hand and right-hand pages alternately. This would avoid the standard unit start and in addition could break through other set patterns which might exist in the presentation of the course. The sets of analysis sheets for each book thus contained as random a sample as could easily be devised.

In addition to the analysis of the first third of each book, the last units in *CEC* and *HI* were also analysed to explore whether any noticeable gradient could be detected in the material, though in the event none appeared. A further analysis was undertaken, on an exploratory basis, of a course being used by the researcher as teacher (*Headway Advanced* Soars & Soars 1989), in connection with the simulations reported on in the next chapter, and another more directly targeted analysis was done of a course being used by teachers who had agreed to take part in the project (*Language in use Intermediate* Doff & Jones 1994). The latter application included analysis of materials written by the teacher: it fed directly into the first structured experiment with a class, which is again reported on in the next chapter. In total, as stated earlier, the consolidating run and the analyses of these currently-used books and materials produced rather over 4500 entries, which seemed an extensive enough application to claim consistency for the system. (The 34 analysis sheets involved, and the further 30 generated by the trial series described in Chapter 6, are available from the researcher.)

In the analysis of all course books, there are consistently 7 tasks entered on Assessment sheet 1 (AS1), designated a to g, and these continue on to Assessment sheet 2 (AS2). This total of 7 was arrived at experimentally as being the number of entries which could be accommodated comfortably on an AS1 and at the same time provide enough variety to cover the progression of tasks within a course unit. At the end of any pair of analysis sheets (AS1 + AS2), the next learner task to be analysed starts five pages on in the student's book. This pattern was repeated for the first third of each book, however many analysis sheets that might require, depending on the book's layout.

The instructions for completing the sheets are given in detail in the Analysis guide, as represented in Appendix 4.1, but in brief, AS1 records what aspects of a task the learner meets in sequence as he deals with it, and AS2 sets levels of difficulty for aspects of the mental processes he needs to go through to complete the task. Further commentary on the development of the system has been given in Chapter 4: the present aim is to go through the procedure step by step to demonstrate how it is put into practice.

5.2 The analysis system applied: AS1

This section explains in detail how Analysis sheet 1 was used to record the content of a course book. The complete set of data for all three course books, as produced by the consolidating run, is extensive; it is also largely repetitive as a procedure, though varied in content. It seems reasonable therefore to illustrate the working of the system which runs from course book through analysis and on to assessment by reporting in detail on the data for one book, and in the following discussion, the analysis of the first third of *Cobuild* 2 is used as a basis. For this demonstration, four samples of material from this part of the course book and the analysis sheets relating to them, as given in Appendix 5.1, will be referred to. Each of these sets consists of the following:

- extracts from the course book, both student's and teacher's pages, which are printed opposite each other in the spirally-bound teacher's book as published. (The student's book is of course also published separately.) It was important for the analysis to consult both the material offered to the learners and the recommendations of the authors on how it should be handled.

(In Appendix 5.1 these eight pages are referenced '1S' to '4S' and '1T' to '4T'.)

- copies of the analysis sheets, AS1 (factual) and AS2 (processing), relating to these extracts from the course book. The entries have been copied from the original pencilled version.

(These eight pages are referenced '1AS1' to '4AS1' and '1AS2' to '4AS2'.)

Excerpts from this Appendix will illustrate how the analysis system is used, and in addition, reference will be made to the Analysis Guide (AG), given in full in Appendix 4.1, to explain how its prescriptions were put into effect.

As a starting point, *Figure 5.1* is a reminder of the headings of AS1, which are set out in 11 columns, and under each of these a note is to be made by the analyser to record his findings after examining SB and TB and taking the advice of AG.

Figure 5.1: AS1: headings

LAC Analysis sheet 1 ref: sheet

task	context				skills		language essentials				processing ⇨
1	2	3	4	5	6	7	8	9	10	11	
	topic	situation / who/where	set	culture	receptive	productive	structure	lexis / field/vocab/func	phonology	discourse / relat/ass/123	

It will be remembered that the analysis takes the learner's point of view and asks questions in the order in which he needs to answer them if he is to do the task set by the course book. The exception to this order is the processing which must occur between input and output (columns 6 receptive and 7 productive), which is allocated an analysis sheet to itself - AS2. (The way in which this sheet is completed will be explained in due course below, after AS1.)

This demonstration of the analysis system begins on page 9 in Cobuild student's book (SB), and the associated teacher's book (TB) page 8; the Analysis sheet is number 51, which occurs part way through the consolidating run. The first row of this AS1 is given in *Figure 5.2*

Figure 5.1: AS1: sheet 51, first row of entries

LAC Analysis sheet 1 ref: Cobuild 2 12, p9 sheet 51

task	context				skills		language essentials				processing ⇨
1	2	3	4	5	6	7	8	9	10	11	
12 languages you've learnt	topic	situation / who/where	set	culture	receptive	productive	structure	lexis / field/vocab/func	phonology	discourse / relat/ass/123	
12a "omit" 12b (a) understand (two Brits talking)	languages	L1 - self / neutral	lone	Brit	(R)L	W	pres simple Q forms	languages / Eng, Fch, Gk, It, get by, struggle / understanding	-	outsider / speaking languages / 1	

NOTE 1:

In the analysis sheet above, each task is allocated a lowercase letter with brackets, whereas *Cobuild* allocates lower case letters alone, that is (a) = analysis; a = *Cobuild*.

NOTE 2:

Quotations in the descriptions of the procedure detailed below come from three sources which are indicated by tinted initials behind the appropriate text;
SB = Student's Book, TB = Teacher's Book R= Recordings.

SB page 9 reads:

12 Languages you've learnt

a What languages can you speak? How many foreign languages have you learnt, or tried to learn? [tape 12b] b Catherine talked to someone called Stephen about the languages they knew. Listen and make a list of the languages they knew, and say how good they were at each.

This text is accompanied by a drawing of several young people in a classroom with speech balloons containing 'Do you speak English?', '¿Se habla español?' and 'Parlez-vous français?' (see Appendix 5.1.1S).

The script of the relevant recording [12b], given on page 135 of TB, is:

SB: How many languages do you speak?
CM: Erm, well, I learnt, erm English, French and German at school, I know a tiny bit of Italian and a tiny bit of Greek, from travels, but I can't say that I really speak any of the foreign languages to any degree, any more.
SB: Well apart from English and American which is fairly similar, I can get by in French and I can struggle through reading Italian and that's about it.

TB comments:

SB12a

1 Don't go into these questions in detail (omit them altogether if you want to do 12g below)....
Listening SB12b [12b](unscripted)
2 Students listen to the tape and take notes about Catherine and Stephen. Suggest they make a table to fill in and give them a framework like the following:

Key	CATH	STEPH
FRENCH	learnt at school	can get by
GERMAN	learnt at school	no
SPANISH	no	no
ITALIAN	tiny bit	reads a little

From these three sources of information (SB, the recorded tape and TB), the title of the unit and entries for the first task may be written in the appropriate columns of AS1. TB says that 12a may be omitted, so 12b is the first task to be analysed.

col 1: task

The defining question for 'task' (as designated under 'Entries (4)' in the Analysis Guide), is: *What do I have to do with this input?* The category 'task' is described in AG as 'closed +', which means entering one of the ten options given (eg 'act out,

complete, discuss, do...understand...') or another considered by the analyser to be more appropriate. In this instance, the students are to listen to a tape and show they have understood it by writing a list related to the content, so 'understand' (one of the given options) is entered. AG also requests that a few words should be added in brackets to explain, and these are entered in the example as '(2 Brits talking)'.

col 2: topic

The entry under topic (defining question: *What are we talking about?*) is 'open', ie not given as prescribed alternatives in AG, but leaving the analyser free to find appropriate words. For this column, a 'short description' is suggested by AG, so in this example, the entry is 'languages'.

col 3: situation - who

AG suggests two options from a closed list of eight to answer the question *Who is involved in this exchange?* and the entry here is 'L1' (=native speaker unedited) with 'self'.

col 3: situation - where

The three options are 'classroom, [topic place], neutral', the first and last closed and the other open, as answers to the question *Where am I when this exchange takes place?* For this task, the entry is 'neutral', glossed in AG as 'place of exchange is irrelevant to topic'.

col 4: set

The defining question is *Who else is involved with me in this exchange?* and the options are 'lone/pair/group/class'. Since the learners are apparently working individually for this task, the entry is 'lone'. Some teachers might do this exercise in pairs, but the TB does not suggest it.

col 5: culture

What do I perceive as the background to this exchange? can be answered by one of six options, or another if needed (closed +): 'British' is one of the options, and is entered here because the speakers on the tape are 'L1 persons' talking of their own experience.

col 6: skills - receptive

How does this input reach me? is clearly by Listening, but there is also a supplementary Reading input in the SB text (going beyond mere instruction or rubric) which explains what is happening on the tape, so the entry is '(R)L'. The R could include guidelines for the list written on the black/white board by the teacher, and this would be coded 'b/board' ('teacher writing for the whole class'), but it is in effect an explanatory rubric, which is to be omitted, as AG explains in the glossary. The drawing in SB might be considered as another supplementary input (vis), but it adds nothing to either L or R texts in terms of understanding, so is ignored for the analysis.

col 7: skills - productive

What kind of output do I have to create to carry out this task? Options are 'S/W/act/draw/tick/none'. SB says 'make a list...' and TB suggests how this should

be done ('...make a table...'), so that the entry is 'W' - 'writing as a productive activity' (AG).

col 8: structure

('the structure(s) which are essential for my completion of this task' - AG) is entered as 'pres simple, Q forms', both of which appear in the taped conversation.

col 9: lexis - field

What is the semantic field within which this task is operating? is answered by 'languages'.

col 9: lexis - vocabulary

As with 'structure', the concept is the minimum essential: 'sample of the minimum few words without which it would be impossible for me to do this task' - AG. The entry lists four names of languages, then 'get by', 'struggle'.

col 9: lexis - functions

What am I doing with these words? is entered as 'understanding' - AG proposes an open list.

col 10: phonology

Every L input and S output involves sound, so that entries in this column are potentially almost infinite. Instead of 'essential' (as with structure and vocabulary), the concept which guides entries in this column is 'features specifically mentioned in the course book as learning points in this task' - AG. *Cobuild* is concerned with phonology only in the context of 'controlled practice' of 'useful and very frequent combinations of words in English in order to build up [students'] confidence and their ability to produce groups of sounds and intonation patterns accurately and spontaneously' (TB:iii). As a result, most entries in this column are blank in this analysis.

col 11: discourse - relationship

What is my relationship with other participants in this task?, with 4 options: 'equal/subordinate/superior/outsider'. In this instance, the entry is 'outsider', since the learner is listening to a conversation without taking part in it - as a 'fly-on-the-wall'.

col 11: discourse - assumptions

What needs to be common knowledge between me and the other participants, to the extent that we do not have to refer to it? An open category, for which the entry here is 'speaking languages'.

col 11: discourse - level of expectation, 1/2/3

The defining question is, *How high are the expectations of my understanding beyond the literal text?*, with 1 low and 3 high. The entry here is 1, in the expectation that learners in an EFL class will have no difficulty in making necessary assumptions about learning languages.

The analysis of the task described in col 1 of sheet 51 is continued with AS2 (on the reverse of AS1), for elements of processing to be entered. But the processing entries are based on a different approach, so it seems best to continue with a report

on a second row of AS1, which begins with a new task.

The definition of a new task for the purposes of the analysis was a problem discussed at some length in Chapter 4. The difficulty lay in deciding what conditions were to be applied to the course book material to indicate when a new task began, and it may be recalled that a set of questions was finally arrived at to guide the analyser in this decision:

1 Is the activity in the student's book?

2 If it is a rubric, is it preparatory (rather than explanatory?)

3 Is it a new topic?

4 Do both 'who' and 'where' change?

5 Does 'group' change?

For the *Cobuild* analysis on sheet 51, the next activity does occur in SB, which is a positive answer to the first question. It reads:

[tape 12c] c Many people say that the British are very lazy about learning foreign languages, because they think that when they go abroad they will usually be able to find someone who speaks English.
To find out if this was true, we asked a variety of British people from different walks of life about their language learning experiences.

TB advises:

Students read the introduction to 12c. Ask them if they feel that English people are lazy about learning other foreign languages. How many British people do they know who can speak their language really well/quite well/just a bit? Encourage them to comment if they wish.

Under 'rubric' in AG glossary, the definition is:

'rubric - Instructions in the course book explaining how to carry out a task.
An explanatory rubric tells you how the task is to be done, and is an integral part of that task (eg Read the story...).
A preparatory rubric introduces topics, structures, vocabulary, etc, and will constitute a task on its own (eg Which of the following can you find in the picture?)'

Under these conditions the introduction labelled 'c' in SB, used as advised by TB, is clearly preparatory, so the second task on AS1 begins with it. In the analysis, working to its own system of labelling tasks from (a) to (g), this task is designated (b).

Figure 5.3: AS1 sheet 51, second row of entries

LAC Analysis sheet 1 ref: Cobuild 2 12, p9 sheet 51

task	context				skills		language essentials			processing ⇨
1	2	3	4	5	6	7	8	9	10	11
12 languages you've learnt	topic	situation / who/where	set	culture	receptive	productive	structure	lexis / field/vocab/func	phonology	discourse / relat/ass/123
12a "omit" 12b (a) understand (two Brits talking)	languages	L1 - self / neutral	lone	Brit	(R)L	W	pres simple Q forms	languages / Eng, Fch, Gk, It, get by, struggle / understanding	-	outsider / speaking languages / 1
(b) discuss (Brits & lang)	ditto	T - students / classroom	class	ditto	(R)L	S	ditto	ditto / ditto, lazy / discussing	-	equal / speaking languages / 1

The script of the relevant recording [12c], given on page 135 of TB, starts:

Meriel West, Travel Agent, currently working in New York.
We were taught French at school, but very badly. We had to memorise lists of verbs, instead of being given sentences to learn which we could use in everyday speaking. I took another French course a few years ago, and we learnt some quite useful things. I also tried to learn Russian, but I did not get very far, although it was interesting. In fact I found a number of words have nearly the same pronunciation in both Russian and English.
Richard H Turner, student of Engineering at Loughborough University.
The foreign languages I was taught at school were Latin, German and French. However, the only language I actually learnt was French. (Although I 'picked up' English at the early age of one.) My French has been particularly useful. For example, reading the instructions on imported packets of French coffee. Also on one occasion, whilst on holiday in France, when a vineyard owner explained how he produced champagne. When in France, the most useful phrase is 'Parlez-vous anglais?' which means 'Do you speak English?'. However I once mistakenly asked a puzzled French man if he spoke French! (He did.)

Two other speakers contribute to the recording, with about the same amount of text, but the first two indicate for present purposes the kind of language and information given. (The whole text is reproduced as transcript in Appendix 5.1.1T, and also as a graphic in SB, as pseudo-real cuttings - from magazines? - irregularly superimposed on illustrations of people and objects which might be connected with the speakers).

TB's advice (above) suggests that students should read the introduction to 12c and then the teacher asks them questions about the British and language learning. This introduction apparently ends at '... language experiences.' in SB, for the next suggestion from TB is about reading and listening (in that order).

So the second task as far as AS1 is concerned is:

col 1: task
(b) 'discuss (British people & languages)'.

col 2: topic
languages, the same as for row 1, so 'ditto' entered.

col 3: situation
teacher discussing with students, a classroom exercise, so the entries read 'T - sts ('when the whole class is participating in the task' - AG), classroom'.

col 4: set
'class'.

col 5: culture
as before, so 'ditto' again.

col 6: receptive
reading the introduction and listening to the teacher's questions and to the other students discussing - '(R)L' again.

col 7: productive
making a contribution to the discussion, so 'S' is entered.

col 8: structure
as before: 'ditto'.

col 9: field
languages 'ditto', vocab as before plus 'lazy', function is 'discussing'.

col 10: phonology
no sign of particular attention to phonology: entry is ' - '.

col 11: relationship
mainly with other students, since teacher is expected only to start off the discussion (and is not regarded by TB as a dominant figure), so this entry is 'equal'; the assumptions are still about speaking languages; and the level of expectation is still '1'.

Two more examples of AS1, as given in Appendix 5.1, have been chosen to illustrate the effects of different kinds of task. The first demands some lateral thinking, involving the invention of a text for a devious postcard. It occurs in the third set of sample pages reproduced in Appendix 5.1 , beginning with page 19 in both SB and TB of Cobuild. The AS1 is sheet 53. It is the second task on this sheet, part of Unit 33 in the course book, entitled 'Holiday postcards'. The relevant part of AS1 is given in *Figure 5.4*. Comments are made on important points arising in the course of the work, and these are added as footnotes, headed ***Observation***.

Figure 5.4: AS1 sheet 53: three tasks

LAC Analysis sheet 1 ref: Cobuild 2 33, p19 sheet 53

task	context				skills		language essentials			processing ⇨
1	2	3	4	5	6	7	8	9	10	11
33 Holiday postcards	topic	situation / who/where	set	culture	receptive	productive	structure	lexis / field/vocab/func	phonology	discourse / relat/ass/123
(a) understand (postcards & reasons for)	postcards	originator - L1 - self - partner classroom	pair, class	WEuropean	RvisL	S	pres, neg, past, modal (cld), fut, pres perf	holidays spectacular, flight, airport, weekend telling news	-	outsider postcard writing 2
(b) invent (a postcard text)	ditto	self - author neutral	lone	own	(R)	W	ditto	ditto ditto + own hinting	-	sub ditto 2
34 [word power] time & money										
(c) do (vocab exercise)	time & money	author - self - partner classroom	pair	Brit	R	tick	past	(uses of) time vocab misc (sentences) [learning vocab]	-	sub/equal learning 1

The first task on this AS1 is not of interest at the moment. It is entered as: (a) 'understand (postcards and reasons for)'; it incorporates the first three activities in the course book. The second activity is therefore labelled 'd' in the course book, but appears as (b) on AS1, the second task as far as the analysis is concerned.

SB reads:

33: holiday postcards

d Imagine you are on holiday abroad. You write a postcard to a friend. Your real purpose is to persuade them to meet you at the airport or station. But of course you want to do this without actually asking them directly. Write the postcard.

TB reads:

Planning

Elicit a few ideas about how to drop hints, eg

We don't get back till 10.30 at night. I'm afraid the children will be very tired.

I hope we don't have to wait too long for a taxi/bus at the airport.

It will be nice to get home but I always hate the long busride from the airport.

I wish we'd left our car in the airport carpark, but it's so expensive we couldn't afford to.

It is probably better not to put these on the board in full, otherwise all you have left is a mechanical copying exercise. In groups students write their postcards. Go round and help as necessary.

In the first task, the students have already read and discussed the content of two postcards reproduced in SB and heard two L1 persons discuss them on tape.

As *Figure 5.4* shows, task (b) on AS1 begins:

col 1: task

What do I have to do with this input? 'Invent' is one of the 10 alternatives given by AG, and the analyser adds the clarification '(a postcard text)'.

col 2: topic

The obvious answer is 'postcards', which is the same as for the previous set of entries.

col 3: situation - who

The demand ('Imagine... Write...') comes from the book, or more precisely, the authors of it, as opposed to the others in the list of closed options - 'course characters, L1 person, originator' ('writer/speaker of text used in course book, as opposed to author = writer of course book'), 'partner, students, teacher'. Entry is therefore 'self - author', since the student is writing to fulfil the authors' request, even though his text is intended for a fictitious third party (see ***Observation 1*** below).

col 3: situation - where

There is no given context, and yet it is not quite a straightforward classroom exercise: entry is 'neutral', a decision buttressed by the fact that postcards can be written anywhere.

col 4: set

TB says 'in groups students write their postcards', but in practice each student is intended to produce an individual version. Entry:'lone' (see ***Observation 2*** below).

col 5: culture

Students are to imagine they are "on holiday", so presumably they are writing home: culture must be 'own' (but see ***Observation 3*** below).

col 6: receptive

Entry is '(R)' because at this point the only reading is of the background to the student's supposed postcard writing.

col 7: productive

The student is to write: entry 'W'.

col 8: structure

The postcard is to be modelled on instances just read and discussed (exercises a - c in SB), so the structures are likely to be any of the same - 'pres, neg, past, modal (could), future, pres perf'.

col 9: lexis - field

As before, 'holidays'.

col 9: lexis - vocabulary

Words as for the task before, since recent use is likely to propose repetition, but the student is also expected to find ways of hinting and may well therefore need to

find his own means, entered as '+ own'.

col 9: lexis - function

Asking without asking is represented by 'hinting'.

col 10: phonology

none.

col 11: discourse - relationship

Author is instructing student what to do: entry is 'sub'.

col 11: discourse - assumptions

What one can or cannot do when 'postcard writing'.

col 11: discourse - expectation 1/2/3

More than the standard 'Wish you were here'; some subtlety is required, but examples are given: entry is '2'.

Observation 1

The use of an imagination-stimulating task produces difficulties for the analysis: although 'author' is the instigator of the exercise, and is in the end the person for whom the postcard is to be written, the intended recipient of it is 'a friend', a kind of role player. But neither SB nor TB characterises the friend (beyond the supposition that he is helpful, and/or gullible), so that the exercise becomes one of devious writing in the abstract rather than a communication with another known person. This is the basis for choosing 'author' as audience.

Observation 2

Another problem with this task is the potential confusion between 'lone' and 'group' in col 4: TB suggests 'in groups students write their postcards' and that the teacher should 'go round and help', implying that the students are to help each other and be helped by the teacher, so that cols 6 and 7 should include (L) and (S) respectively. This may be taken as an admission by the authors that the task is too difficult for students working on their own, yet the final product seems intended to be a postcard from each individual. Somewhat unforgivingly, 'lone' is entered in col 4 and the skills remain (R) and W alone.

Observation 3

The location of the friend to whom the postcard is addressed is not given, but the assumption must be that the writer is returning (home) from holiday. In that case, why not write in L1?

The next task on the same AS1 is concerned with a different topic (which defines a new task in accordance with rule 3 - see above). It is one of a series of 'wordpower' sections recurring throughout the course book, intended to 'cover particularly common target words which have a number of important meanings or

uses, for example the word **way**' (TB:iv).

SB says:

Wordpower: time and money.

1 time = minutes, hours, days, week, months etc.

Mummy and Daddy went back to Ireland some time ago.

It'll rain all the time.

How did they spend the time?

2 to have a _____time.

It sounds like you had a good time.

We had a dreadful time.

3 time = an occasion when something happens

The second time was when we went to eat at a hamburger restaurant.

I've already boarded this flight five times and every time I ended up in Cuba.

This time our friend left a warning note.

4 time = two o'clock, 8,30, 1720 hours etc.

What time is it? Do you have the time?

5 times

The taxi cost five times as much as the bus.

Which category do these examples belong to?

(a) Took a very long time getting there.

(b) People we met on it said they'd been many times before

(c) Look at the time. We're going to be late.

(d) But next time I knew I could do it.

(e) We had a very busy time at work last week.

(f) My job is a hundred times more difficult than playing the piano.

[then another exercise with 'spend money/time'].

TB advises:

1 In pairs students match sentences a - f to the correct lexical categories. Refer them to the Lexicon for further explanation of the categories.

2 Students find the words in sentences g - l [in the second exercise] which can be used with both time and money.

[Key to exercise given]

3 Ask students if there are words in their own language which can be used with both time and money. In English there is a saying: 'Time is money'. Do they have a similar expression?

AS1 therefore is completed as follows:

col 1 task

This is a vocabulary exercise based on sentences which are unconnected with each other except by particular words that they contain. The entry in the first col is

therefore 'do' ('mechanical exercise which has no social context'- AG) with '(vocab exercise)' in explanation.

col 2: topic

The same as the title of the section in SB: 'time and money'.

col 3: situation - who

TB says students are to work in pairs, and this presumably carries through to the second exercise, but as with task (a) above, the origin of the material is the text-book writers, so the entry combines these participants in the entry 'author - self - partner'.

col 3: situation - where

In this case, clearly a classroom exercise: entry is 'classroom'.

col 4: set

In accordance with TB, entry is 'pair'.

col 5: culture

The exercise seems intended to relate to CHE ('common human experience of a cultural kind, eg Do you get up as soon as you wake up?' AG). But in spite of the AmE-influenced neologism 'It sounds like (sic) you had a good time', the mixing of simple past and present perfect with 'already' in ' I've already boarded... every time I ended up...' and references to hamburgers and Cuba, the idiom and ambience are essentially English, in an almost Christopher Robin vein: 'Mummy and Daddy... a hundred times more difficult... how much did you actually spend?' Entry is therefore 'British'.

col 6: receptive

The exercise is generated in print in SB: entry is 'R'.

col 7: productive

The end product of the exercise is matching sentences with categories, so the entry is 'tick' ('marking non-verbally to indicate a choice ... in tasks requiring ... numbering, lettering, underlining etc' AG). There is also some L and S involved because the work is to be done in pairs, but this is not the object of the exercise (see ***Observation 4*** below).

col 8: structure

The sentences are mainly in past tenses, so entry is simply 'past'.

col 9: lexis - field

The point is vocabulary learning, mainly about time, so the entry is '(uses of) time, vocabulary'.

col 9: lexis - vocabulary

Here is a case where the salient words could be listed, but in fact there is much more to the sentences than merely 'time' and 'spend'. The entry is an admission of defeat: 'misc (sentences)' (see ***Observation 5*** below).

col 9: lexis - function

The use of square brackets round '[learning vocabulary]' is intended to indicate that the object of the exercise is not a functional achievement but a learning activity.

col 10: phonology

None again: entry is ' - '.

col 11: discourse - relationship

The authors are in charge, but pairs discuss, so entry is 'sub/equal'.

col 11 discourse - assumptions

The purpose of the exercise, as students are aware, is as entry says: 'learning'.

col 11 discourse - expectation 1/2/3

For the students, learning is the point of being in the classroom: entry is '1'.

Observation 4

This task illustrates the problems of differentiating in the analysis between an activity intended to improve learner's skills in exchanging meanings in order to achieve an end (communicating) and an exercise which takes individual items of language with little or no context attached and asks learners to learn them. It introduces into the decision-making the question as to whether all skills involved should be recorded, even if peripheral to the main object of the task, as in the use of pair work in a vocabulary exercise. The difficulty is that class work of whatever kind may be done in pairs, so that the entry 'pair' in col 4 set could become universal and therefore undifferentiatingly meaningless. The fact that TB suggests it does not mean that the exercise is necessarily intended for the promotion of learners' discussion techniques. In practice, the attempt was made to enter 'pair' or 'group' only when this method of working promoted exchange of opinion on a topic external to the classroom, and not when it was used as a collaborative effort to find answers to preordained questions. This is a fine line to draw, but seems necessary in the context of defining variations in task demands.

Observation 5

A second problem exemplified by this task is the recording of vocabulary in col 9. The shortcomings of word lists as prescribed learning aims, and their even more detrimental effect on test specifications, were discussed in Chapter 3; for the analysis, the need for alternatives to listing is acute, partly because of limited space in AS1, but more fundamentally because if learners are to make the best of what they know to reach the conclusion to a task, they need to be given the freedom to use whatever vocabulary they can retrieve, from whatever source. They are not restricted by the need to 'know' a stated body of content, as sometimes demanded by teachers to examinations (Page 1983); the analysis has indicated areas of topic and lexis within which (and beyond which) the learners are enabled to operate by the relevance of the scenario.

5.3 The analysis system applied: AS2

An account of the implementation of AS2 is now required to complement the one that has just been given of AS1. The same four tasks will be used to exemplify the working of 'processing' as recorded by the analysis system, but the reference will now be mainly to AG, since SB and TB have already been extensively quoted above in the explanation of AS1. (It will be recalled that the Analysis Guide is given complete in Appendix 4.1, and the relevant pages of Analysis sheets and course book material appear in Appendix 5.1.)

AS2's gradings of processing elements are made under 8 sub-headings: input, response, load, search, link, know, individual and demand, and below these at the head of AS2 are given the highest available grade for each: a denominator which varies from /7 to /3. As explained in Chapter 4, decisions are to be made in relation to defining questions in AG (as with AS1), and in addition there are keywords for each concept, to help focus the analyser's understanding of the field intended. Again, examples are given in AG, but this time they cover the whole range of judgements for each concept: they were all taken as examples from entries on AS2 in the consolidating run. The analysis again takes the learner's point of view, so that the questions asked (for example, 'How dense is the input and how difficult is it to take in?' - sub-heading 12) have a learner's perspective. This puts into practice Gipps' suggestion (1994), which was referred to in Chapter 3, that one way to promote consistency is to set up criteria and illustrate them with exemplars. In addition, it includes reminders (in the form of keywords) of the concepts on which the criteria are centred.

Figure 5.5: AS2 sheet 51: rows 1 and 2

LAC Analysis Sheet 2: processing

11	12	13	14	15	16	17	18	19
task (as col 1)	input	response	load	search	link	know	individual	demand
	/6	/5	/7	/6	/6	/6	/3	/6
(a) understand (2 Brits talking)	2 simple text 9 lines, L	3 W list of names	2 L2 x L1s 9 lines	1 L & list	1 recognise names of languages	1 names of languages	2 re speaking languages	1 W list of 8
(b) discuss (Brits & languages)	2 R,T talk, sts discuss	3 S if choose to	3 class discussion	2 recall own experience	1 people & my language	1 own experience	1 own experience	4 S to class

Task (a) is 'understand' (2 Brits talking)', which is carried across from AS1. The entries are now made under sub-headings rather than under columns as on AS1.

sub-heading 12: input /6

- defining question: *how dense is the input and how difficult is it to take in?* (AG).
- keywords: density, quality (AG)

Learner listens to a tape of 2 L1 persons, 9 lines of text in the TB (set in a column 7.5 cm wide). (see ***Observation 6*** below)

- examples of gradings: high 6: understand joke, R TV listings & discuss...; low 1: L sounds for stories, vis pics for 'going to'... (AG)
- grade entered: 2, with comment 'simple text 9 lines, L'.

sub-heading 13: response /5

- defining question: *how quick/psychologically pressurised a response is expected from me?*
- keywords: pressure, speed, competition

Learner lists languages and names and checks the one against the other on the basis of notes made from listening to tape.

- examples: high 5: L numbers & W, speedread & S to class...; low 1: in own time, homework...
- grade entered: 3, 'W list of names'.

sub-heading 14: load /7

- defining question: *how many inputs are there and how many are happening at once?*
- keywords: quantity, simultaneous, N skills

Learner listens to 2 Brits in turn.

- examples: high 7: 30 lines of transcript, R 700 words, all class involved...; low 1: sounds, R grammar point, L partner...
- grade entered: 2, 'L 2xL1s, 9 lines'.

sub-heading 15: search /6

- defining question: *how much logic/imagination/memory do I need to carry out this task?*
- keywords: find, invent, work out, remember

Learner lists languages and names and checks the one against the other on the basis of notes made from listening to tape.

- examples: high ó: improvise, formulate narrative, work out pattern from examples...; low 1: fit to pattern, produce in parallel, remember own experience.
- grade entered: 1, 'L & list'.

sub-heading 16: link /6

- defining question: *how difficult is it for me to transfer word to idea, text to meaning, & vice versa?*
- keywords: apply, extract, compare, rearrange

Learner extracts information from tape and applies it in list.

- examples: high 6: extract discussion & S, apply principle to this case...; low 1: apply greetings to situation, apply learnt rule...
- grade entered: 1, 'recognise names of languages'.

sub-heading 17: know /6

- defining question: *how far is the task/content/ situation likely to be familiar to me?*
- keywords: know, recognise, connect

Learner hears names of languages.

- examples: high 6: specialist topic, principles of organisation...; low 1: own convictions, own experience, facts about self...
- grade entered: 1, 'names of languages'.

sub-heading 18: individual /3

- defining question: *how committed am I likely to be to the task, how relevant is the content to me?*
- keywords: own, interest, useful, intriguing

Learner lists languages and names from tape.

- examples: high 3: hypothetical, no necessary identification, no relevance...; low 1: own convictions, own experience, facts about self... (see **Observation 7**)
- grade entered: 2, 're speaking languages'.

sub-heading 19: demand /6

- defining question: *what is expected in my response?*
- keywords: fluent/accurate, vague/exact, closed/open

Learner lists languages and names from tape.

- examples: high 6: accuracy essential, fluent & accurate, L & guess what...; low 1: fluency not expected, underline grammar examples in text, discuss if want...
- grade entered: 1, 'list of 8'.

Observation 6

The transcript related to task (a) is 77 words long, but it was decided not to make word-counts of listening or reading texts for the analysis, to save time.

Observation 7

The grading for sub-heading 18 'individual' is in an odd way the reverse of all the others in the cline of difficulty: for other sub-headings, the top grade represents difficulty and the lowest, least demand. For 'individual', the top grade represents the frustrations of boredom and the lowest, the ease generated by enthusiasm.

The next task to be discussed is the second on AS2 (see *Figure 5.5* above) and as in column 1 of AS1 is headed:

(b) discuss (Brits & languages)

sub-heading 12: input /6

- defining question: *how dense is the input and how difficult is it to take in?* (AG).
- keywords: density, quality (AG)

Learner reads statement about British being lazy about learning languages, then

discusses reactions with teacher and other learners.
- examples: high 6: understand joke, R TV listings & discuss...; low 1: L sounds for stories, vis pics for 'going to'... (AG)
- grade entered: 2, 'R, T talk, sts discuss'.

sub-heading 13: response /5
- defining question: *how quick/psychologically pressurised a response is expected from me?*
- keywords: pressure, speed, competition
Learner expected to contribute to class discussion, but no competition implied.
- examples: high 5: L numbers & W, speedread & S to class...; low 1: In own time, homework...
- grade entered: 3, 'S if choose to'.

sub-heading 14: load /7
- defining question: *how many inputs are there and how many are happening at once?*
- keywords: quantity, simultaneous, N skills
Learner has short introductory R text, listens to teacher, talks with rest of class.
- examples: high 7: 30 lines of transcript, R 700 words, all class involved...; low 1: sounds, R grammar point, L partner...
- grade entered: 3, 'class discussion'.

sub-heading 15: search /6
- defining question: *how much logic/imagination/memory do I need to carry out this task?*
- keywords: find, invent, work out, remember
Learner compares own experience with other learners' views.
- examples: high 6: improvise, formulate narrative, work out pattern from examples...; low 1: fit to pattern, produce in parallel, remember own experience.
- grade entered: 2, 'recall own experience'.

sub-heading 16: link /6
- defining question: *how difficult is it for me to transfer word to idea, text to meaning, & vice versa?*
- keywords: apply, extract, compare, rearrange
Learner hears and expresses views on Brits' language learning.
- examples: high 6: extract discussion & S, apply principle to this case...; low 1: apply greetings to situation, apply learnt rule...
- grade entered: 1, 'people & my language'.

sub-heading 17: know /6
- defining question: *how far is the task/content/ situation likely to be familiar to me?*
- keywords: know, recognise, connect
Learner listens/tells experiences of Brits' language learning.
- examples: high 6: specialist topic, principles of organisation...; low 1: own

convictions, own experience, facts about self...
- grade entered: 1, 'own experience'.

sub-heading 18: individual /3

- defining question: *how committed am I likely to be to the task, how relevant is the content to me?*
- keywords: own, interest, useful, intriguing

Learner will have tales to tell, compare.

- examples: high 3: hypothetical, no necessary identification, no relevance...; low 1: own convictions, own experience, facts about self...
- grade entered: 1, 'own experience'.

sub-heading 19: demand /6

- defining question: *what is expected in my response?*
- keywords: fluent/accurate, vague/exact, closed/open

Learner will be expected to join class discussion.

- examples: high 6: accuracy essential, fluent & accurate, L & guess what...; low 1: fluency not expected, underline grammar examples in text, discuss if want...
- grade entered: 4, 'S to class'.

This commentary is repetitive, which suggests that after some practice, analysers are likely to find the operation of the grading system fairly straightforward. The explanations for the other two examples of AS2, following on from the commentary given earlier on AS1 sheet 53, will therefore be presented in more abbreviated form, without defining questions, keywords and examples, which have already been set out twice. As with AS1, the first task on AS2 is to be omitted for present purposes.

Figure 5.6: AS2 sheet 53: rows 1 - 3

LAC Analysis Sheet 2: processing

11	12	13	14	15	16	17	18	19
task (as col 1)	input	response	load	search	link	know	individual	demand
	/6	/5	/7	/6	/6	/6	/3	/6
(a) understand (postcards & reasons)	4 R 2 postcards	3 discuss implications, pair, class	2 two postcards + discussions	2 implications of 2 postcard contents	4 postcards' implications	2 travel, holidays	1 intriguing: why?	2 discuss partner
(b) invent (postcard text)	1 rubric alone	2 W: follow text for model	1 1 input (rubric)	5 find story, persuade	4 express ideas in W	1 own story	1 own story	4 W postcard
(c) do (vocab exercise)	2 examples of 'time'	2 study, tick	4 11 sentences, match with 4 concepts	2 apply rule	2 match sentences to concepts	2 CHE	3 just boring, bits	1 tick

The second task on AS2 begins (see *Figure 5.6*):

(b) invent (postcard text)

sub-heading 12: input /6
Learner is asked to write devious postcard to friend.
- grade: 1, 'rubric alone'.

sub-heading 13: response /5
Learner writes postcard in collaboration with partner, little pressure, model text available.
- grade: 2, 'W: follow text for model'.

sub-heading 14: load /7
Learner to read 6 short lines of instruction, consult with partner.
- grade: 1, 'one input (rubric)'

sub-heading 15: search /6
Learner needs to remember models, invent means of indirect persuasion.
- grade: 5, 'find story, persuade'.

sub-heading 16: link /6
Learner needs to formulate ideas indirectly in writing.
- grade: 4, 'express ideas in W'.

sub-heading 17: know /6
Learner writes own ideas.
- grade: 1, 'own story'

sub-heading 18: individual /3
Learner's own ideas.
- grade: 1, 'own story'.

sub-heading 19: demand /6
Learner required to write postcard, accuracy desirable, not essential.
- grade: 4, 'W postcard'.

The interest of this task is in its demand for originality: the learner is expected to find a way of requesting help by hinting that it is required, but not asking directly. This is reflected in high gradings for *search*, *link* and *demand* - that is, finding the means with which to express what is to be expressed - while the task mechanics (*input*, *load*) and the content (*know*, *individual*) rate relatively low.

The third task on AS2 begins (see *Figure 5.6*):

(c) do (vocab exercise)

sub-heading 12: input /6
Learner is given examples of the uses of the words 'time' and 'spend'.
- grade: 2, 'examples of "time"'.

sub-heading 13: response /5
Learner discusses possibilities with partner and matches new sentences to demonstrated categories.
- grade: 2, 'study, tick'.

sub-heading 14: load /7
Learner reads 11 sentences demonstrating 4 different uses of 'time', considers 6 more sentences for matching, considers 7 quasi-synonyms for 'spend'.
- grade: 4, '11 sentences, match with 4 concepts'

sub-heading 15: search /6
Learner is to find concepts exemplified in sentences.
- grade: 2, 'apply rule'.

sub-heading 16: link /6
Learner is expected to compare new sentences with given examples.
- grade: 2, 'match sentences to concepts'.

sub-heading 17: know /6
Learner is expected to find situations (in which sentences are set) to be familiar.
- grade: 1, 'CHE' (= common human experience).

sub-heading 18: individual /3
Learner reads disconnected sentences for vocabulary exercise.
- grade: 3, 'just boring, bits'.

sub-heading 19: demand /6
Learner required to match new sentences with exemplified concepts.
- grade: 1 'tick'.

This task contrasts with the preceding one: it is expecting not creativity but a literal application of rules presented in sample sentences. It expects high levels of simultaneity, the carrying across of ideas from one place to another, but contains relatively low grading everywhere else, except in levels of potential boredom, which are high.

These four examples of tasks have shown how the analysis system is applied to the course, with reference to both student's book and teacher's book (with transcripts of listening input), including extracts from the Analysis Guide, which explains how the system works and offers examples of entries and definitions of terms used. This excursion through the completion procedures has covered the mechanics of Assessment sheets 1 (mainly facts) and 2 (mainly judgements) and has shown how the system works in practice - cumbersome to begin with, but becoming easier to apply as the method is learnt by the analyser. At the same time it is evidence that the system is a consistent mechanism, asking questions which are relevant to the learner's situation and providing answers on a principled basis which builds up into a substantial system of consistent judgement. The next section can now deal with how the data accumulating from the 4500 entries mentioned earlier - the result of the consolidating run - were used as a basis for a specification to guide the writing of assessment materials.

5.4 The data: Assessment Sheet 1

The results of the analysis as recorded in AS1 are to be grouped according to the use which can be made of them. (Since the entries for AS2 are made on a different basis, they will be discussed separately.) The first step in interpreting the AS data is to add up the occurrences of entries for each column. This was done by totalling entries at the bottom of each AS1, then carrying these totals forward to a grand total which, as explained earlier, represents the analysis of the first third of the course book. These totals were then grouped into three sections, to provide various kinds of information.

Section 1: *Quantity* simply adds up the number of occurrences, and is applied to the closed categories (ie those with alternatives specified in the Analysis Guide) in cols *1 task, 3 situation, 6 receptive* and *7 productive*. The ratings as reported in *Table 5.1* are derived from the number of tasks analysed for that book, 7 per analysis sheet, with 4 - 6 sheets per book, the number of sheets depending on the amount covered in the first third. For *Cobuild*, 4 sheets x 7 tasks gives a factor of 28, so the task labelled as 'discuss', occurring on 10 out of 28 possible occasions, is rated 10/28 = .36. (This system incidentally puts ratings for all course books on the same basis, enabling comparisons to be made between them.) *Table 5.1* lists the ratings under the column/headings of AS1 and adds an explanatory note for each, derived from the analysis guide in Appendix 4.1.

Table 5.1: Summary of AS1 ratings for *Cobuild* 2, pp9-25: interactions (quantities)

column/heading	*grading*	*notes (from glossary, Appendix 4.1, Analysis guide)*
1 task		
complete	.	fill gaps in text
discuss	.36	discuss topic/exercise with others
do	.21	carry out instructions for mechanical exercise
find	.07	find information, structures, words...
invent	.	invent (eg complete a story)
tell	.07	tell, narrate
understand	.28	show understanding of text
3 sitn: who		
author	.29	writer of course book as participant in task
L1	.29	native speaker of English, unedited
originator	.18	writer/speaker of text used in course book
partner/s	.57	other learner(s) in pair/group
st/s	.25	student(s) when whole class is participating
T	.14	teacher as participant in exchange
3 sitn: where		
classroom	.79	place where task occurs if not topic place
neutral	.25	when place of exchange is irrelevant to topic
7 receptive		
L	.61	listening as a receptive activity
R	.61	reading as a receptive activity
vis	.14	visual representation as part of input
b/board	.	black/white board used by T for whole class
(R)	.32	reading as subsidiary to other skill(s)
(vis)	.	visual as subsidiary to other skill(s)
8 productive		
S	.68	speaking as a productive activity
W	.25	writing as a productive activity
(S)	.11	speaking as subsidiary to other skill(s)
tick	.07	marking non-verbally to indicate a choice
none	.07	where no entry can be made

<.05 is represented as .

The most frequent kind of *task* is shown to be 'discuss' (.36), followed by 'understand' (.28) and 'do' (.21). As a consequence, *partner* provides the most important input (.57), far ahead of 'author' (.29) - which is associated with contextless exercises (eg on grammar) - and 'L1'(.29). This last figure, relating to the use of native speakers both as input for understanding and as models for discussion/discovery activities, is surprisingly low, considering the authors' emphasis on 'real English' as supplied by the foundation corpus. The *situation* is most often 'classroom' (.79), which implies little role play activity, for then the entries are made in inverted commas, as 'airport', 'interview room', 'on phone'. (These examples are from *CEC*.)

Section 2: *Coverage* takes note of the range of material included, and is applied to the open categories in cols *2 topic*, *9.1 field* and *11.1 assumptions*.

Table 5.2: Summary of AS1 ratings for *Cobuild* 2, pp9-25: coverage (range of material)

column/heading	*N*	*notes (from glossary, Appendix 4.1, Analysis guide)*
2 topic		
languages	6	what are we talking about?
jobs & pay	4	
grammar	3	
lexis	2	
postcards	2	
holiday	2	
9.1 field		what is the semantic field in which this task is operating?
languages	7	
jobs & pay	5	
personal	5	
holidays	4	
vocabulary	2	
11.2 assumptions		what needs to be common knowledge between me and the other participants, to the extent that we do not have to refer to it?
speaking languages	6	
personal description	4	
jobs/pay	4	
learning	3	
postcard writing	2	
holiday brochure	2	
jokes	2	

Table 5.2 contains tallies of the number of times entries appeared under these three headings which, with the overlaps between them, show the areas of interest considered appropriate by the authors of the course books for the intended learners. The overlap suggests that grouping topics under wider areas might be a way to summarise topic choice for the specification and hence for the assessments. A review of all three course book analyses indicates that *people*, *society*, *travel*, *money* and *tradition* might be helpful megagroupings of this kind, but then it could be argued that justifications can be found for fitting in virtually any topic under these five. The problem is that selecting topics on the basis of those found in course books is either too restricting because too specific (and also therefore suspect in terms of generalisation), or too random because grouping of topics tends to be so general that the result is no longer a helpful reflection of what has been discussed in previous learning. One solution is to link the entries into themes, eg *language-travel-holiday-postcard-description*, so that there is coherence of topic as well as more likelihood of demonstrating transfer to other situations. This idea is taken up again in Chapter 6, in the context of materials writing.

Section 3: Location is concerned with how far the course includes different interactions and contexts, and is based mainly on cols *4 set*, and *5 culture*, with references back to aspects of *11.1 relationships* and *11.2 assumptions*. The results of the analysis for *Cobuild* appear in *Table 5.3*, where again notes from the Analysis Guide explain the headings.

Table 5.3: Summary of AS1 ratings for *Cobuild* 2, pp9-25: location (quantities)

column/heading	*grading*	*notes (from glossary, Appendix 4.1, Analysis guide)*
4 set		
lone	.39	when learner is working on own
pair	.46	when one other learner participates in task
group	.14	more than pair but fewer than whole class
class	.43	whole class involved together in task
5 culture		
own	.18	culture of learner's own background
CHE	.11	common human experience, personal
British	.54	culture of text or task is British
WEuropean	.	culture of text or task is West European
worldwide	.18	culture as worldwide, not personal
US	.	culture of text or task is North American
11.1 relationship		
equal	.68	learner equal with other participant(s) in task
sub	.36	learner subordinate to other participant(s)
outsider	.11	learner not directly addressed (fly on the wall)
11.3 difficulty of assumptions		
1	.61	low expectations of understanding beyond text
2	.25	mid ...
3	.18	high ...
<.05 is represented as		

The data for *location*, as for *interactions* (*Table 5.1*), are presented as proportions. *Cobuild* apparently uses *pair* work most (.46), closely followed by *class* (.43), then *lone* (.39), with *group* work well down (.14). This reflects the pattern of a typical *Cobuild* unit which has learners working alone or with a partner to consider some kind of input, then reporting back to the class as a whole, rather than working in groups. The main background culture is *British* (.54), with *own* and *worldwide* both at .18. Reflecting the high incidence of pair and class work, the most important relationship is *equal* (.68), with some *sub* (.36) as the result of exercises instigated by *author*. Finally, there seems to be little difficulty for learners in understanding the implications of the settings for interaction, the lowest level of *assumptions* scoring .61.

5.5 The data: Assessment sheet 2

An entirely different kind of data is provided by AS2, which is summarised in a fourth section of data grouping.

Section 4: *Complexity* is concerned with processing. Whereas the other three sections report mainly on factual elements in the course book by adding up instances as they appear, this section represents findings from the use of AS2, which attempts to deal with the cerebral activities postulated as requirements for fulfilling the tasks listed. The data appear in the form of gradings, as demonstrated in the examples worked through above. The listing in *Table 5.4* below shows the median score of all grades allocated under each sub-heading, representing a set of typical values of possibly widely fluctuating gradings made on relatively short ranges of possible scores. The maximum grade for each sub-heading (ranging from /6 to /3) is also given. The explanation in each case is an adaptation of the defining question used in judging grades.

Table 5.4: Summary of AS2 grades for *Cobuild* 2, pp9-25: processing (medians)

sub-heading	*median grade/maximum*	*notes (from glossary, Appendix 4.1, Analysis guide)*
input	2/6	how dense is the input & how difficult is it to take in?
response	2/5	how pressurised a response is expected from me?
load	3/7	how many inputs & how many are happening at once?
search	3/6	how much imagination/logic/memory do I need?
link	3/6	how difficult is it to transfer word to idea & vice versa?
know	3/6	how far is task/content/situation familiar to me?
individual	2/3	how committed am I likely to be to the task?
demand	3/6	what is expected in my response?

The value of this exercise depends on how these variations in difficulty (from 2/6, = 33% to 2/3, = 66%) are to be reflected in the writing of assessment materials, a formidable demand which appears at first sight difficult to fulfil, unless a vague middlingness of difficulty is to be aimed at throughout. At this point it seems necessary to explore further the purpose and potential of the AS2 analysis.

As discussed at length in Chapter 4, the theory was that if links could be systematically established between the cerebration required for effective language use and the actual event, it might be possible to devise means of assessing how well learners were exploiting whatever language they had available in the conditions set by a task. This theory includes a range of different approaches to description which run from abstract concepts resting on statistical justification, for example the many competencies proposed by Bachman (1990:87) to practical detail, for example the

marking guide set out in ARELS AP29, which explains how to decide what is creditable in a learner's answer and what credit to give for it (ARELS nd:14ff). The definitions arrived at in the development of AS2 lie at some mid-point along this scale from abstract design to building brick.

Skehan's (1998) account of the interplay between task and assessment has a different perspective, though it grows from the same root, since the sub-headings of AS2 were based on his categories (see Chapter 4 above). But he develops the questioning of what goes on between input and output by investigating firstly the influence of learners' adaptation to the conditions under which tasks are done and secondly the varying characteristics of the tasks themselves as they affect performance. His list of task conditions which have been shown experimentally to affect learners' performance can be linked back to the sub-headings of AS2, as *Table 5.5* shows.

Table 5.5: Comparison of task difficulty factors (Skehan 1998) and AS2 sub-headings

small number of elements	load
concrete information	link (+ input?)
immediate, here-and-now	search (+ individual?)
retrieval of information	search
familiar information	know

This comparison shows that the work done by Skehan & Foster (1997) confirms the AS2 categories as being significant in the assessment of the output from task-based interchanges, but brings the interpretation of the results of AS analysis no further forward. More generally, the present project takes into account factors which Skehan (1998:176) considers influential, for example allowing the learners time and opportunity for control by setting open-ended tasks, reducing the 'stakes' by engendering an informal setting and engineering surprise not only by unexpected events but also by introducing unlikely contexts for discussion. But these are considerations which appeared during the development of materials, which is to be dealt with in Chapter 6.

For the present, a working hypothesis suggests that further ways of deriving information from the analysis provided by AS2 could include:

- investigating horizontal patterns in the AS2 results rather than vertical accumulations and so producing a profile of the demands of a particular task: this would allow investigation of relationships which might exist between a task type and its demands under particular sub-headings;

- using the analysis system on assessment materials in draft, to see how far they match the outcomes of the prior analysis of the course book;

- comparing learner performance on a task written to fulfil an analysis with that same analysis so as to explore whether the demands of the assessment task have indeed met what the analysis system proposed as a specification.

These are more general indications of the impact of processing on learner performance and task content, but are still derived from the interaction of the learner and the material he is to cope with, in an area which is central to understanding of how the influence of mind and attitude affect language learning in general and especially the individual learner's responses.

5.6 The data: Comparisons between course books

An important result of the analysis of the three main course books in the consolidating run is the possibility of comparing one with another to establish whether the system differentiates between them, and if so, what differences it suggests. Since the analyses of *Cambridge English Course (CEC)* and *Headway Intermediate (HI)* were summarised in the same way as *Cobuild*, comparisons are justifiable and could be instructive. The equivalent data for all three course books are laid out below in *Tables 5.6 - 5.9*. A commentary follows on each set of data.

Under the heading of 'interactions' (*Table 5.6* - cf *Table 5.1* above), *Cobuild* and *HI* prefer tasks which expect students to 'discuss' and 'understand', and this implies group and pair work and concern with input. The highest rating for *CEC* is for 'do', with a wider range of other task types such as 'discuss', 'guess', 'understand' and 'match'. *Cobuild* and *HI* have a more limited range of task types, concentrating on three in each case. There is little difference between the courses on how far the tasks are directed by 'author', but *Cobuild* uses 'L1' speakers to the same extent as 'author', whereas the others use them not just less, but not at all. *Cobuild* is shown to be by far the most frequent user of work with partner(s); *Cobuild* and *HI* are more likely than *CEC* to involve the whole classful of students, but *CEC* uses more role play, judging by the five occasions on which participants appear in inverted commas, to indicate a 'let's pretend' situation. As regards skills, *HI* is most concerned with listening, and the other two are not far behind. *Cobuild* gives equal weight to listening and reading, whereas the other two give reading less, but still substantial, emphasis, and visuals are less important in *Cobuild* than the other two. On the productive side, speaking is by far the most important skill demanded by all three, but especially in *HI*; writing is required more in *CEC* than in the others, but not by much.

Table 5.6: Summary of AS1 ratings for three course books interactions (quantities)

column/heading	*Cob*	*CEC*	*HI*
1 task			
act out		.	.06
complete	.	.08	.
discuss	.36	.15	.34
do	.21	.21	.
find	.07	.08	.11
give own		.	
guess		.13	.06
invent	.	.	.
tell	.07	.05	.05
understand	.28	.18	.23
learn		.05	
match		.13	.
recall		.	
summarise			.
3 sitn: who			
author	.29	.28	.29
course ch/s		.	
L1	.29		
originator	.18	.13	.20
partner/s	.57	.21	.37
st/s	.25	.23	.31
T	.14	.18	.11
'advertiser'		.	
'applicant'		.	
'interviewee/r'		.	
'officialdom'		.	
'tourist'			
		.	
3 sitn: where			
classroom	.79	.77	.63
neutral	.25	.13	.26
'airport		.	
'home'			.
homework		.	
'in same house'			.
'interview room'		.	
'neutral'			.06
'on phone'		.08	
7 receptive			
L	.61	.64	.71
R	.61	.49	.49
v	.14	.21	.20
bb	.	.	
(L)		.05	.09
(R)	.32	.10	.26
(v)	.	.	.09
8 productive			
S	.68	.72	.80
W	.25	.31	.23
(S)	.11	.	.
(W)		.08	.
act			
draw		.	.
tick	.07		.
none	.07	.01	.

<.05 is represented as

Table 5.7: Summary of AS1 ratings for three course books: coverage (range of material)

	Cob	*CEC*	*HI*
2 topic			
languages	6		
jobs & pay	4	4	
grammar	3	8	1
lexis	2		
postcards	2		
holiday	2		
descrip/people		6	1
jungle escape		3	
future		2	
travel		2	
phon		2	
food			5
daily life			3
TV			3
children/sport			3
pop singer			3
women			2
11.2 assumptions			
speaking langs	6		
personal descrip	4		
jobs/pay	4		
learning	3		
postcard writing	2		
holiday brochure	2		
jokes	2		
mechanical		8	
letters		4	
descriptions		4	
jungle		3	
future		2	
tourism		2	
phoning		2	
stereotypes			7
food			5
daily life			3
TV			3
sport & children			3
holidays			2
9.1 field			
languages	7		
jobs & pay	5		
personal	5		
holidays	4		
vocab	2		
descrip/people		6	
jobs		4	
grammar	3		
future	3		
superstitions	2		
stereotypes			5
food			5
stereotypes			4
sport			3
TV			3
holidays			3
advertisements			2
news			2

The 'coverage' ratings for *CEC* and *HI* in *Table 5.7* show that the correspondence of topics among the three course books is extremely thin, 'jobs & pay' and 'descriptions of people' being the only two common areas among the 17 topics listed. This spread of topics could indicate either that there are so many possibilities that authors' choices are not likely to coincide, or that the sampling of the course books is not dense enough (for example, that other coincidences may exist but do not appear within these samples), or that the one- or two-word summaries for the analysis do not reflect potential overlaps (for example, postcards, holiday, travel might all have been counted as aspects of the same area for discussion).

Table 5.8: Summary of AS1 ratings for three course books: location, quantities

	Cob	*CEC*	*HI*
4 set			
lone	.39	.31	.20
pair	.46	.08	.31
group	.14	.44	.11
class	.43	.46	.60
5 culture			
own	.18	.08	.11
local		.	
CHE	.11	.18	.06
British	.54	.18	.40
WEuropean	.	.13	.17
worldwide	.18	.18	.31
US	.	.05	
none		.18	.
11.1 relationship			
equal	.68	.33	.60
sub	.36	.59	.37
sup		.08	
outsider	.11	.05	.
11.3 difficulty of assumptions			
1	.61	.77	.89
2	.25	.21	.11
3	.18	.	

<.05 is represented as .

Table 5.9: Summary of AS2 grades for three course books: processing (medians)

	Cob	CEC	HI
input /6	2	4	3
response /5	2	3	3
load /7	3	4	4
search /6	3	4	4
link /6	3	4	4
know /6	3	3	4
individual /3	2	2	2
demand /6	3	4	4

Table 5.8 is concerned with 'location', and a comparison of the three course books shows that they have different patterns of grouping: under 'set', *Cobuild* is shown to concentrate on pair, class and lone, in that order; *CEC* on class, group and lone; and *HI* on class (by far the highest level), pair and lone. The reason for *HI* appearing so

frequently class-oriented may be that this course offers less specific teacher guidance than the others, suggesting an authors' assumption that it is for the individual teacher to decide what set is most appropriate for each task. *Cobuild* has a predominantly British cultural context, whereas *CEC* has equal commitment to *CHE* (common human experience), British, worldwide and none, and *HI* is mainly concerned with British and, to a lesser extent, worldwide. West European figures in both *CEC* and *HI*, but not at any great frequency. The main difference in 'relationship' is between *CEC* and the others: it places the learner in a subordinate role more than half the time, with *Cobuild* and *HI* rating 'equal' in about two-thirds of cases. Finally, the degree of difficulty for the learner in entering into the assumptions made by the speaker/writer is higher for *Cobuild* than the others.

Under 'processing', in *Table 5.9*, it appears that *Cobuild* sets lower levels of complexity than the others, *CEC* demanding most for 'input' and *HI* most for 'know'.

In brief, *Cobuild* is characterised by the use of native speakers as models, derived from the Cobuild corpus but also from the results of asking native speakers to attempt the tasks set for learners. This input is used for pair work, which reaches higher levels here than for the other two books. There is also more class participation in discussions. *CEC* on the other hand seems relatively authoritarian, more often demanding direct learning of grammar points and lexical areas, though relying on work in groups as well as for the class as a whole. For *HI*, the important points are that listening and speaking are more frequent as skills than reading and writing, and that working as a class is more frequent than in pairs and groups.

These comments on course books, made possible by the analysis system, are a by-product of a development concerned with the impact of course book material on learners. But the information is useful in various ways because it investigates learning material with a perspective different from that taken by other textbook evaluation procedures (reviewed in Chapter 4). It investigates what demands are made on learners and so more easily distinguishes author assumptions and styles, indicating for example the effect of policies such as reliance on LI speakers and a corpus of English in use; the relative importance given to different language skills; and relative expectations in the areas of mental processing. Possible uses extend from appraisal of new materials to investigation of a course's use of particular components, for example listening as against reading or formal study as against group interaction.

5.7 The analysis as a system

The development work on course book analysis has resulted in a principled system which produces useful information about the material presented to learners. The approach is empirical: it takes the course book as found and analyses it from the learner's point of view, which means that by asking what the learner is to do with the tasks presented, a detailed account of teaching/learning material is built up.

This provides comprehensive information about the diverse demands of the course book, but resolves it into comprehensible components in a sequenced pattern of activities. The analysis is based on two patterns: factual and judgemental. The first investigates what the learner needs to do and know to fulfil the demands of the task set by the course book: it is justified by the logic of the learning events, taken in order. The second pattern, which relies on informed judgements about the relative difficulties in the demands made by processing, is justified by a theoretical structure which is derived from well-documented debates in the literature.

The system is valid to the extent that it reflects what happens in the course book and brings this forward as a basis for assessment; that it accepts the theoretical stance of the course book authors towards language learning; and that it is a convincing translation of what learners and teachers do in class into a comprehensible coding procedure. It was developed through a series of experiments with a range of course books, using iterative applications of a trial system with successive amendments until stability was reached, then a consolidating run to prove its applicability. This extended development programme has resulted in a sophisticated analysis procedure which studies in detail the interactions between learner and material, and learner and learner. The stability achieved and the range of material analysed means that the system is not only specific to given series of events in one particular course book but also universal: it can be used for the analysis of language learning activities in general. It is therefore consistent - or reliable - because it is founded on extended trials, follows a logical sequence which is transparent for analysers and so likely to produce similar results both over a period and in comparison with other analysers, and can be applied as a standard system to any language learning material. The outcome of all this development work is practical both because its content relates directly to factual events in the classroom and because its computer-based operation makes it relatively quick to learn and use in relation to the amount of detailed information it provides.

5.8 Organising the specification

Providing for the individual

The most familiar use of the term 'specification' relates to the description of components in constructions such as buildings or ships or engines; within examination boards it has connotations of precise prescriptions of what shall appear in tests. Munby uses it (1978) for his 'sociolinguistic model for defining the content of purpose-specific language programmes'. The difficulty here is that many language learners (for example those in the early stages at secondary schools) have no goal defined for them beyond a general awareness of how words and things are conceived differently in a foreign country compared with at home. In language learning for those with definite purposes in view, the specific aspects of a specification refer to their needs for language use in professional or working life,

such as the case of the Austrian waiter offered as a paradigm in early Council of Europe papers (eg Bung 1973). The problems begin as the specification turns from a prototypical Austrian waiter to one particular man living in a particular area, dealing with a particular type of customer. Munby's characterisation is so specific that if it is to be taken literally, his model can operate only with named participants, that is 'an individual or the stereotype of a category of participant' (Munby 1978:52). When Carroll comes to apply Munby's procedure to putative overseas students aiming to attend courses at British universities, he sets up examples of participants to illustrate a range of possibilities. However, these are not the result of 'comprehensive observational studies of the participants actually engaged on their courses', but profiles compiled by subject-qualified British Council staff, which were then personalised so as to 'focus the collection and interpretation of data on a real, or at least putative, individual' (Carroll 1981:69-71). Clapham's comment that 'without such [observational] studies, surely the profiles are almost useless' (Clapham 1981:113) may be worth making, but a more serious principle is at stake. In the wider perspective, the multiplicity of detail built into the profile of one individual for the purposes of test specification is reduced in value in proportion to the extent to which each candidate finds the test less than specific to his individual needs. Tests are not designed for individuals, they are designed for groups: but within this generality individuals may be given scope to define themselves. For the individual, validity lies not in matching the test with his own uniqueness or even his own typicality, but in how far the testing system reflects the situations he will meet and the part he will have to play in them.

This is the basis for the present approach to specification. The course book has been sampled to provide a pattern of language encounters as made by students and the analysis thus represents what they have experienced in class. The project now needs to encapsulate that experience to provide the basis for an assessment.

Use of data

The summary of the analysis of *Cobuild* is an example of what the system produces, but it is not yet in a form which can be used directly by a writer of assessment material to guide him in his choice of content. The summary generally follows the order suggested by the learner's point of view, so that *task* comes first, followed by *situation* and *skills*. The writer needs a different perspective, since his development work has its own logic. For example there is a need to set up as a starting point a plausible situation in which the learners' language is to be deployed, and then to consider what tasks are likely to occur in that situation. A helpful approach is therefore to start from *topic*, since this will lead on to consideration of the context in which the learners are at this moment, so that the assessment tasks they deal with may fulfil the main aim of the project: individual feedback to learners with relevance to their situation. After topic, the next consideration is the *culture* within which the action is to take place, followed by the *tasks* to be devised, the set in

which they are to be done and the *skills* required to fulfil them, with *structure* and *lexis* to be found as required for topic and task. The *discourse* elements will relate to topic and set. Finally, the *processing* needs are to be considered in relation to the preceding decisions. This succession of elements suggests a linear approach to writing, but in practice there is interplay between one element and another, with adjustments being made continually as the idea for an assessment exercise develops.

The condensation of the analysis data into a specification - in effect, a brief for a materials writer - is illustrated in *Figure 5.7* with, as before, the data from *Cobuild*.

Figure 5.7: Specification

Cobuild 2: Units 1 - 4, pp 1 - 29

Topic: languages I have learnt, relationships with others, holidays and postcards, jobs and pay

Location: British

Tasks: understand, discuss, do

Set: pair, class, lone

Skills: R, L; S

Structures: present, past, comparatives

Lexis: [as required by topic]

Relationship: equal

Assumptions: [as required by topic], Difficulty: 1

Processing: input 2/6 (=33%), response 2/5 (40), load 3/7 (43), search 3/6 (50), link 3/6 (50), know 3/6 (50), individual 2/3 (66), demand 3/6 (50)

This specification is intended to be used as a guide, not as a cast-iron framework into which assessment materials are to be made to fit. Although efforts were made to sample the course book in a way which can be defended as representative, the result is inevitably an interpretation of the content, and it seems unreasonable to use it rigidly to control what will eventually be an open-ended assessment. The aim is to base the assessment on the previous work without demanding exact correspondence with the results of the analysis procedure. For example, the *Cobuild* analysis rates *group* at 14%, compared with 46% for *pair* (see *Table 5.3*), which would imply that the assessment materials should not include much work for groups. On the other hand, from the point of view of assessment, there is a need for group working, to give learners the opportunity to hear and respond to a variety of points of view and to contribute to a discussion, which in *Cobuild* is evidently (according to the sample analysed) carried out in pairs or with the whole class together. Pair work is useful for sorting out ideas but is not interactive enough to represent varied discussion; class work is useful for coming to a consensus but may put undue pressure on the individual who is not happy to talk in front of others. Group work, though figuring small in the analysis, should therefore be given

a rather larger share in the assessment materials because it can provide evidence about collaboration and a participant's awareness of other participants. The important principles here are that the specification is the foundation for the assessment and that all of it is covered, but excursions outside it, if promoting the common sense of the tasks set (another way of looking at 'authenticity') and exploration of a wider range of learner skills, are to be included.

Information on context

An important principle, established as a result of the review of testing systems reported in Chapter 2, is that assessment material should relate directly to the learner's situation. Some way is needed of establishing what this means, and the obvious answer is information from the school, and more specifically from the teacher of the class concerned. The procedure followed in the present case was to conduct preliminary guided interviews with teachers who were to participate in trials of the system, and at the same time to collect non-instructional material which was available to students at the school, such as leaflets about events both in the school and outside it and on noticeboards. This information was intended to clarify the content of class work, characterise the students and suggest plausible events to which the assessment material could be related. The interviews with teachers were based on the questionnaire reproduced in *Figure 5.8*.

Figure 5.8: Enabling questionnaire

School ...	*class ...*	*teacher ...*
1. What book is this class using?		
2. What use is made of teacher's book (if any)?		
3. Are any other books used?	courses	
4.	resource books	
5.	other	
6. What material provided by the school?		
7. What material provided by teacher?		
8. What individual learning provided?	self-access	
9.	computers	
10.	other	
11. Are there any options for groups/individuals?		
12. What are the students like?	ages, interests (ESP)	
13.	nationalities	
14.	how long been here	
15. What extra curricular activities?	trips	
16.	culture classes	
17.	sport	
18.	other	
19. Where is the school?	town	
20.	area	
21.	things to see/do	
22.	with families/accomm?	
23. Other		

The way in which all this information is gathered together and used as the basis for the writing of assessment material is the subject of the next chapter, which gives an account of how preliminary experiments developed into a series of trials in one school enabling both course book and situation to be represented in the assessment materials.

5.9 Summary

This chapter has demonstrated how the analysis system was put into effect for the sample of *Cobuild* used in the consolidating run, representing the first third of the course. Four samples of the analysis have been pursued in detail, with references to three sources of information: the student's book, the teacher's book (including transcripts of tape recordings) and the Analysis Guide. This showed how each of the elements contributed to the implementation of Analysis sheets 1 and 2 as they were applied to the tasks set by the course book. The resulting data were collected into groups which related to different parts of the analysis and were then condensed into a form which could be used as the basis for a specification. A detour was made from this sequence from data to specification for an investigation of what the analysis showed of the similarities and differences between the three course books used for the consolidating run of the analysis system, and some conclusions drawn about the varying approaches of the three sets of authors.

Finally, a further source of input to the specification was described: a questionnaire with which teachers were consulted to enable the context of the learners to be characterised so that assessment tasks could be designed which related to their current situation.

The next chapter is concerned with the development of assessment materials and the conduct of trials with learners in schools.

Chapter 6

Devising and trialling scenarios

6.0 Introduction

After the demonstration of the analysis system and of the way in which a specification is derived from it, the next step is a report on the development of assessment materials and the trials undertaken with them in schools. The materials are another link in the chain from the investigation of ideas and the formulation of principles to practical applications in classrooms. The first section of this chapter is concerned with experiments in designing formats for exercises which capture the learner's interest and at the same time reflect course book content. Theoretical principles and practical possibilities are discussed in turn, the one informing the other. This exploration starts from the perspective of the learner's personal involvement, proposing situations for learners in Britain which are built on a sequence of realistic commitments such as finding somewhere to live while attending an EFL school. This leads on to an account of further experimental work with materials written to exploit the background of the learners' situation as found, first in Perú, then at a school local to the researcher and finally at a school in London. The enabling questionnaire, used to guide an interview with the class teacher ('enabling' because it made possible the writing of pertinent material, and so named to distinguish it from the 'self-assessment' questionnaire used later with students), directed the content towards realistic tasks for learners in a particular place. It was introduced as part of an early trial at a London school, where the run of the sequence from introduction to analysis to assessment materials was first used, to investigate the viability of the system so far.

The result of this development was a system for designing an activity in a format which was at the same time specific to a situation and yet flexible enough to accommodate the demands of a preceding analysis. This activity was labelled 'scenario', and the second section of this chapter deals with this concept and its organisation. A definition of the term is followed by guidelines for writing scenarios, with comments on their format and content.

A series of trials is reported in the third section. The trials follow the sequence of procedures described in previous chapters, from course book analysis to scenario. They were undertaken in one school over a period of three months, and show what progress the project has made in practice. The data resulting from the trials are then discussed in a fourth section.

6.1 Development of assessment materials: preliminary research

Early proposals

The analysis had resulted in guidelines for some kind of interaction, controlled by the content of the preceding course. But at the same time, the response of participants was to be uncontrolled, for any appropriate language they used and any appropriate conclusions they reached were to be accepted. The search was therefore for some set of mechanisms which would

- relate to past learning;
- offer open-ended choice to participants, both in decision-making and language use;
- relate to the present, that is the current social and geographical contexts of the participants;
- relate to the future, in that it would provide for feedback to participants on their potential performance outside the classroom;
- include activities which would enable participants to show what they could do with their available language, both individually and in cooperation with others.

Examples of activities which fulfilled some of these aims could be found in simulations and other problem-solving activities, as discussed in Chapter 3, and from them could be drawn principles which would inspire participants with O'Neill's (1987) 'motivational push', for example, working through conflicting evidence towards agreement (Lynch 1977, Jones 1983), taking part in progressive groupings towards a whole-class decision (Allwright 1979, Brims 1982) and engaging the perspectives, opinions and choices of individuals (Maley, Duff & Grellet 1980). But these ideas needed to be integrated with guidelines laid down by the specification, representing both the content and the motivations of the course book.

Another Working Paper (WP4) was written at this point to explore possibilities for the writing of assessment materials. Starting from the importance given to the individual learner in the work so far, the first ideas were principally concerned with where the learner was placed, that is, his current situation. The discussion then expanded to include meeting the requirements proposed by a specification. An early proposal was to start from student experiences in Britain, as in the CAE (University of London 1986) which in turn had been derived in part from the ARELS oral examinations (1967). (These test systems were described in Chapter 2.) Any town in Britain of a reasonable size, if set in an area of interest and containing one or more schools of EFL, could be the starting point for the invention of a fictionalised but realistic town (for example, a combination of Cheltenham, Gloucester, Bath etc to be called 'Chelster') which could serve for any L2 student

in Britain as a place for a tourist visit or for learning English. For source material, 14 collections of realia on various topics were supplemented by photographs taken by the researcher of daily happenings in Cheltenham, some of them intriguing (a window cleaner with ladder on bicycle, odd characters, people chatting, street scenes, eccentric shop fronts), and a recording of a local person talking about the town with the help of prompts from some of the material collected.

A series of situations was set up, for example an invitation from an English contact for the learner and a friend to spend a weekend in Chelster, and group of tasks connected with attending a local EFL school, including looking for somewhere to live. But after extensive planning of drafts, the Chelster enterprise seemed capable of producing only tasks which were either touristic or boring, or both; they lacked the immediacy required for students to feel personally involved in events and decisions. Several principles did however emerge from these explorations. Among them are the following.

Fictionalisation. Invention frees the materials writer from the literal, allowing him to add variety and the unexpected to stimulate interest, while still enabling him to represent fundamental 'truth'. Realism is not lost: authors from Hardy to Chatwin fictionalise the places as well as the events they write about, and this step away from reality enhances rather than diminishes the universality of their tales. At the same time, even though for Hardy Dorchester is Casterbridge, London remains London; with Chatwin the happenings on the Black Hill are invented, but the Great War has a considerable impact on them. Local events become generalised, national events remain historical. Novelists are skilled at showing us the essential truth of their characters and events, and materials writers concerned to stimulate the self-expression of learners could seek to provide opportunities for similar universal human truths to be explored in language classrooms. The aim should be to encourage learners to make something new out of the existing old. This means that even if the stimulus to language is not strictly 'authentic' (as discussed in earlier chapters), the response can be. This is an instance of 'considering authenticity not as a quality residing in instances of language but as a quality which is bestowed upon them, created by the response of the receiver' (Widdowson 1979:165) - and also, in the present case, fed by the imagination of the speaker. When he is successful, the materials writer offers a task which is motivating in itself and therefore draws necessary language from the learner - if the necessary demand to perform is there, he can be confident that the performance will come.

Choice. The learner must be allowed to make his own decisions and deal with the consequences. This sets the writer problems of coordination and follow-up, since plausible alternatives must be available, yet controlled by overlapping tracks, as in a maze exercise for language learners (eg Doff & Jones 1994:22ff), so that there is no commitment to writing an exponentially increasing set of alternatives. Choices made between advertisements, for example, cannot be allowed to result in ten

diverging sequences of story with the need for materials to be invented to match each of them.

Variety. On the other hand, alternatives are necessary. They can include the same (or similar) information but given a variety of slants. For example, the accommodation advertisements which were written for the Chelster exercise played variations on the same basic elements, and the listening texts which followed represented two different landladies' introductions to essentially similar accommodation offered. Alternatives reinforce a situation's conditions, save time and space, and promote factual and social contrasts.

Sequence. It is helpful both to writer and learner if the sequence is presented in a standard format - to the writer because it reduces the load of invention (and explanation in rubrics) and to the learner because it makes it easier for him to understand what is expected, especially after experience with the principles of a given format. The idea of sequence can apply to set (lone/pair/group/class) and activities (listen/speak-&-listen/write/speak-&-listen/report verbally/vote/write...) as well as the events of a storyline. Familiarity and clear, logical connections are helpful in many ways.

The next step in the development was the use of experimental materials in teacher workshops and in classrooms, where the current context of the work could supply the immediacy lacking in the Chelster exercises.

Experimental materials in the field

An early example of an interactive exercise which was intended to form the starting point for the assessment of learner's speaking skills was written by the researcher as an illustration for discussion at a workshop with teachers in Perú. The courses being used in the teachers' classes were *Cambridge English course 1* and 2 (Swan & Walter 1985), but no detailed analysis of them was undertaken, for the material was intended as an experimental example of a possible assessment exercise, written to generate discussion at the workshop. The exercise was based on a letter which purported to have been received by the Director of the school, asking for suggestions for activities to occupy the writer's teenage son, who was accompanying him/her on a business trip to Lima. The teacher was to read the letter to the class, then have the students work in groups with publicity material about local events and finally agree a programme for the visitor. The last stage was for each student to write a letter to the visitor explaining what was proposed, and the best letter would actually be sent. The situation was one which could occur in actuality (similar letters had in fact been received by the Director) and it involved students directly in a sequence of events which were realistic and required local conditions to fulfil it; it enabled the students to be themselves rather than play roles; it gave the students something realistic to aim for, with a slight edge of competition to it; and it set up a logical sequence of tasks, with both listening and reading as inputs and speaking and writing as outcomes.

The participants at the workshop wrote materials for this exercise, and some were interested enough in the idea to try it out with students in their classes between sessions. Recordings and transcripts were made of some of the students' spoken texts and examples of their letters were also brought to the workshop for discussion.

The main achievement of this exercise as an assessment mechanism was that although it depended on information specific to its context (classes of adolescents in Perú), the *principles underlying it could be adapted for universal application.* This was an important development which was to be become one of the basic principles on which later material was devised.

Building on the experience of the work in Perú, the researcher wrote experimental materials for his own EFL classes, which he called 'simulations', mainly because they did not follow a detailed analysis of the course book in use. Three of these were written: *Travel award, Alton Towers* and *Books for the library.* The first two of these are now to be described, but *Books for the library* was used again with little alteration in the trial series, and therefore appears in Appendix 6.2 with the other trial materials.

The *Travel award* simulation consisted of a set of interviews: one group of students were interviewers and another were interviewees. (A similar exercise appears in the *Bellcrest Series*, OUP 1973.) The interviewees were to work in pairs to make an application for an award which allowed two people to spend four weeks anywhere in the world for any purpose they were interested in, for example Third World support projects. They were given guidelines for writing a short statement and were then to be interviewed. The interviewers were to plan the proceedings and then conduct the interviews, eventually having to report recommendations to the awarding body (ie the class as a whole). Both groups were given useful phrases as starting points.

The visit to *Alton Towers* began with a newspaper article about this venue which all students were to read. The students were then given guidelines on how the visit was to be planned, by three groups of students, sorting out among themselves within each group which individuals would undertake the various necessary tasks, for example finding out details of the events available, working out a timetable, arranging transport, organising food and so on. Finally, the class came together, the three groups presented their programmes and all together then decided which was the best.

These materials were more experiments in task work for students than exercises set up for assessment purposes, but were intended to be both useful work for the students and helpful experience for the writer. The students reported that they found them interesting and instructive and had enjoyed the experience; the main learning points for the writer were in organisation, such as fitting a number of tasks into the time available, working out a viable sequence of 'sets' - lone, pair, group, class - and working at intelligibility of instructions. All the classes were using *Headway Advanced*, but no attempt was made to relate the content of the simulations directly to the course book: the aim was to involve the students in

discussion which would demand practical use of the language they knew. 'The involvement every participant feels in the tasks they have to do means that for a few hours a simulation can seem just like real life, where everyone is responsible for their own decisions and actions and shares in any collective decisions taken, too.' (Jones 1983:1).

The third development was undertaken at a school in London. The aim was to carry out the proposed scheme in full for the first time : an analysis of the materials used in class was to be followed by a specification, and a scenario (defined below) written on the result. The London school offered a class for a trial, with the promise of sophisticated recording equipment to provide the means for a review of students' spoken texts after the event. The outcome of the trial was intended to consist of:

- the results of the enabling questionnaire, used for the first time as a basis for an interview with the class teacher;
- a sequence of Analysis sheets relating to the course book and the teacher's own material;
- the scenario, consisting of introduction sheet for the facilitator/animateur, a note sheet for each student, texts of postcards and challenge cards;
- extracts from student written responses;
- answers to the self-assessment questionnaire completed by students at the end;
- a tape recording of students' spoken contributions to the discussion;
- a report on the results of the trial.

The enabling questionnaire brought out one relevant and potentially disturbing fact at the start: the alternative teacher of the same class used a course book - *Language in use, intermediate* (Doff & Jones 1994) - but the teacher to be involved in the trial mainly used material she had herself prepared. This was a potential hazard for the system, for if the analysis is to be carried out in advance, the teacher-made material needs to be integrated with the use of the course book (if any), so that the sampling system can be applied for the analysis. This means that the teacher must be able to predict the order in which his own materials will be used, so that they can be treated as, or in parallel with, a course book. In this trial, the teacher's class work for the period leading up to the scenario was well planned and so could be given to the researcher at an early stage for analysis in the usual way. There was no difficulty in applying the analysis system to this original material because the learner's question - What am I doing here? - was still valid and the system was capable of answering it within the categories resulting from the development work with course books. But integration is essential: for example, the use of a set of teacher-made materials without reference to a planned sequence - or at least not one made available to the researcher - vitiated an attempt at collaboration at another school.

The responses to the enabling questionnaire also showed that the students had considerable autonomy in the work they did beyond the core class work, with a wide range of choice, that they were adult (18+), of mixed nationalities, and took part in local sports. Their main interests in the environment of the school seemed to be to exploit the possibilities of being in London, at all levels of cultural sophistication.

The analysis was applied to both *Language in use* and the teacher's material, sampled as before. Two outline scenarios were sent to the teacher so that she could choose which she thought would be more appropriate for the class, and the preferred scenario was then written. The trial was to occupy an hour's class time, and a recording was to be made of the interactions resulting from the scenario. All this was agreed in advance, and the analysis (10 days) and scenario writing (5 days) were done within the time limits set.

The scenario was *Holiday acquaintances*, an early version of *Compatible friends*, which was used in the trial series later (see Appendix 6.1). It began with a pool of 40 postcards, of which each student was to choose one which reminded him or her of a holiday they had had and then talk about it with a partner. The second task was to discuss in groups of four handwritten postcards which threatened a visit from someone met previously on holiday, with a variety of personalities represented and the decisions resting with students as to whether they wished to see the acquaintance again or not. As an additional element in the ensuing 'telephone call', unexpected interpolations ('challenge cards') were handed out to students for them to derail the discussion and provoke unplanned responses. This strategy encourages learners to use their available language creatively to solve problems rather than rely on phrases learnt as cliché responses to all likely situations. The scenario was written as instructions for a teacher/animateur, with set times suggested for each phase so as to fit the hour available, with a note sheet for each student on which he was to make notes as instructed, to guide the later exchanges.

At the end of the scenario, students were asked to fill in a questionnaire with sections on self-assessment and on how successful they thought the exercise had been. This was again the first use of this part of the sequence, and the results were somewhat equivocal. Some students asked for clarification, and this indicates that the questions were not necessarily understood and therefore accurately answered. The questionnaire was revised for future use.

The school had supplied a state-of-the-art omnidirectional microphone attached to a high quality recorder. This combination was exceedingly sensitive and picked up any voice speaking in the room. It was thus ideal for recording individuals but could not cope with three groups of students all talking at once: the tape contained an uninterpretable mush of voices which made a transcript impossible.

A report was written to summarise the results of the trial. There were several positive outcomes. The exercise stimulated the students into talking about topics that were familiar from the preceding classroom work, and they responded with

interest to the situation they were given. There was more writing in this scenario than in later ones, and some students showed both skill and imagination in this part of the sequence. The researcher made considerable progress in the development of the scheme, especially in finding ways of identifying students, accommodating various aspects of timing and undergoing the negative experience of using a recording technology too advanced for the circumstances.

These three experiments were developmental stages in arriving at a viable system for the project. They provided opportunities for the application of the analysis, trials of the enabling and self-assessment questionnaires, development of the scenario as a vehicle for learners' participation and experience in tape recording, all of which were important preparation for the series of trials which was to follow.

6.2 The scenario

Somewhere in the course of the deliberations about materials writing, encompassing the arguments in WP4 and the researcher's debates with teachers, 'assessment unit' became 'scenario'. The change represents a different attitude to the materials. Units imply too close an identification with the compartmentalisation of the course book, coming at stated intervals, seeming to set finite and predicted ends. The discussion which ran in parallel with the development of materials suggested that what is needed above all is an assessment which connects directly with where the learner is now, at this moment, in this place; is open-ended enough to allow him to use his judgement in making choices; and so will lead him on into discussion with other learners and eventually to decisions. For the purposes of this project, this concatenation of factual, affective and linguistic events is called a 'scenario'.

Connotations of the term 'scenario' include (formally) the outline of a dramatic work and (informally) the setting for a joke. The meaning given by Chambers (Schwartz et al 1988) for its informal use (annotated *'loosely'*) is 'any imagined, suggested or projected sequence of events...' But Di Pietro's definition for language learning purposes is more restricted: 'A scenario is a strategic interplay of roles functioning to fulfill personal agendas with a shared context' (1987:41). This definition gives priority to the exercise as a controlling agency rather than to the 'personal agendas', an interpretation supported by the activities Di Pietro quotes as examples of the exercise. For him, a scenario sets up situations with roles beginning: 'You are...', 'You have just...', 'You must...' and continuing with instructions as to what participants' circumstances and attitudes are to be, laying down the parts to be played. In sum, the principles are the same as those of what is commonly regarded (in course books and supplementary materials) as a role play, where the learner is not 'I' but someone else. As used for the present project, a scenario has more in common with a simulation (Jones 1982), where each participant brings his own character and ideas to bear on a problem which he is to discuss with others and possibly solve, even if only reaching a potential solution to a potential problem, eg that of finding somewhere to live. The factors in common between Di Pietro's role plays,

Jones' simulations and the present scenarios are that conclusions resulting from the discussion are not preordained, and that the participants use whatever language they have available to interact with others in order to reach their own conclusion.

According to the criteria as finally formulated for the present project, scenarios are:

- *plausible*, asking learners to work at tasks which are realistic for them in their present context;
- *external*, representing a demand for some useful application outside the class;
- *local*, relating to some event which is appropriate for the learners at that particular school; and
- *personal*, involving each learner in discussions and decisions which depend on his own character and understanding.

Even though as a structure a scenario now appears to have an independent life of its own, it is required to put into practice the specification which has been derived from the course book analysis. This means that the writer needs to find some way of integrating the specification, the scenario's demands for different kinds of interaction and the circumstances surrounding the learners. The information given in response to the enabling questionnaire for teachers now shows its value: what the learners are to talk about is essentially local - in place, in time and in the experience of participants. The tendency is therefore towards the individual in all senses - each scenario represents a unique occasion, matched with existing circumstances. This makes it all the more important for a common layout and sequence to be set up so that the writer has a structure to fill rather than having to invent one afresh each time. This is not to say that scenarios must follow a rigid format, since the demands of a situation may require variations, but in principle it is simpler for all concerned (learners, teacher, writer) if there is a straightforward and familiar pattern to the activity.

Over the period of development, the pattern of a scenario eventually stabilised as a connected series of tasks which followed a standard sequence, though with variations as necessary for particular topics. The standard sequence was:

1 a reading task (also incidentally or largely visual) (Rv) for learners working individually, with written checks (W); then -

2 a comparison/discussion for a pair (SL), with a further written check on the outcome (W);

3 a group discussion (SLR) in preparation for -

4 a class discussion/consensus (SLR).

The different interpretations of this sequence are shown in the report of the trials in the following section. The assumption is that all the skills required of the learner

by this sequence have appeared in the course book and hence in the analysis, and that discussions of various kinds - pair, group, class - will give enough scope for learners to use their available language in ways which the analysis has captured. The the examples devised for the trial series proved adequate to fulfil these conditions, as will be reported in Chapter 7.

The essential elements of an operational scenario can now be characterised from two points of view, content and administration, with an added note on the learner's perspective.

The content of a scenario is to:

- be built on the specification derived from the course book;
- depend on a direct link with the learners' physical context;
- consist of a sequence of tasks following the logic of an overall purpose;
- incorporate all four of the standard language skills (LSRW) in realistic combinations, plus others (eg visual recognition) as required;
- follow a sequence of 'sets' - individual, pair, group, class;
- propose a conclusion to be reached by the whole class together.

Advance planning is important, since three aspects of timing need to be taken into account. When an analysis is to be undertaken of a course book in current use, time needs to be allocated for it to be done in parallel with the class as it proceeds, with predictions of the date when the assessment is to take place and of the point in the course book which will have been reached by then. The second aspect to advance planning is how much class time is to be allowed for the scenario (a typical allocation has been one 45-minute lesson), with the consequent judgement on how much student activity the scenario can demand within the time available. The last timing consideration is how soon the results can be given back to the learners, and this depends on the procedure chosen, from class discussion immediately afterwards to more considered judgements by the teacher in the form of marks or scales or statements, and by the learners in the form of self-assessment procedures. (These outcomes are considered in detail in Chapter 7).

Another aspect to administration for the trials was identification of participants. It was obviously important to know which student was which in the sound recordings if a transcript was to be made as a basis for the consideration of marking schemes. A system was worked out over the series which ensured clear identification in nearly all cases. The situation is of course easier for a class teacher, who has a good chance of recognising voices. Briefly, the triple system eventually used in the trials was: to write a reference on each learner's note sheet, identifying the group and his physical position in it; to make a note on the marksheet of some particularly loud or characteristic phrase as it was said, for linking later to the recording; and in the

last resort to ask the teacher to listen and identify. The use of video recording would have made learner identification easier and would have added a new dimension to the assessment by the possibility of observing, and perhaps even assessing, further aspects of discourse, such as body language and what anthropologists know as gesturality (eg Giglioli 1972). This was considered but not pursued, on the basis that it would add further complication to an already complicated assessment procedure.

For the conduct of the trial, it was also important for the researcher

- to allow time and space for discussion with the teacher both before and after;
- to set aside time for the course book analysis and the writing of the scenario before the trial;
- to write a post mortem report for the record immediately after the event.

From the learner's point of view a scenario should be seen as a useful way of spending a lesson, with helpful feedback resulting from it. Elements which will help to ensure this positive reaction are:

- the conviction of his teacher that it is worth doing;
- some kind of introduction which explains what a scenario is and what it does, including essentially that it is up to the individual learner to act and make decisions within the framework set up for him;
- clear indications of what he is expected to do (rubric);
- enjoyment of the activity while it is going on and a feeling of achievement at the end; and finally,
- useful feedback information about what the learners have done both individually and as a class.

This section has reviewed the development of the scenario as a means of providing a focus for student interactions. The idea of a simulation, which involves learners in fictitious events which lie outside the classroom, has evolved into a more directly situation-based activity which takes into account the students' current here-and-now and so provides motivation and relevance for their use of the language at their disposal. At the same time, a scenario has cogent rules which enable it to be designed and written according to the demands of a preceding specification, derived in this instance from learning material used in class.

6.3 Trial series

A school in Bournemouth agreed to help with the trial series, and did so at regular intervals over a period of three months.

The main information from the enabling questionnaire, which was completed at the

beginning of the series, but later amended to take account of new course books, included:

- course books: *Work Out Upper Intermediate* (Radley & Millichip 1993), supplemented with teacher's materials; *Headway Advanced* (Soars & Soars 1989) and *CAE Advantage* (Kingsbury et al 1992) were the course books for later trials;
- a wide range of ancillary resources in the school;
- mixed European and Far East nationalities, aged 14-20+;
- regular student visits to London and places of interest all over the West of England;
- local interests general;
- all students accommodated with families.

This information made several contributions to the design of scenarios. Most importantly, the course books were the basis for analysis and specification, as the project demanded. But the questionnaire was useful in many other directions. For example, the scenario *Introduction to Dorset* starts from the fact that the students have been in the area for some time and so would know what places might be of interest to new students; *A job in Dorset* includes advertisements for jobs suitable for the age group; and *Compatible friends* exploits the idea of holiday postcards, familiar from their situation in a holiday resort and the temporary nature of their stay there, with the likelihood of transient friendships.

Over a period of three months, the analysis-to-scenario procedure was repeated seven times. The first six scenarios were for two classes taught by the same teacher in succeeding lessons. For the first four, alternate units of the same course book were being used for each class. The next two were for classes using different course books, and the final one was an examination preparation class. A report was written immediately after each trial.

The analysis system worked well and produced specifications within the time available for the first four trials, but time was short for the fifth, so an abridged version of the analysis was used. This was an enforced but useful development, since the system had come to seem cumbersome with repeated use, and overlaps in categories had been suggested by the data, in which repetitions began to appear over time. The chief difficulty remaining in the annotation itself was in column 9 *vocabulary*, for no easily applicable solution had been found to the problem discussed earlier of how to record the 'essential' vocabulary for the completion of a task. The solution used was to refer to the broader category of *lexis*, including *field* and *function*, so that learners would not be expected to 'know' specifically listed words, but could exercise retrieval skills such as inferring and guessing in the receptive activities and simplification and approximation in production. In both cases this means that learners would need to employ strategic approaches to the

exchange of meanings so as to exploit whatever language they had available. This decision has wide implications, and these are taken up again in Chapter 7.

The number of students in each class varied from 12 to 8, and in groups, from 4 to 2. In total, there were 69 student participations in the 7 scenarios, some students taking part in more than one, but none in all. The introduction to the students, made verbally by the researcher and/or the teacher, described the exercise as an experiment to see how far the scenario related to the work done for the course book, with no mention of 'test' or 'assessment' beyond the self-assessment questionnaire at the end. The reason for this was that students were not intended to be pressured by external demands, but only by their interest and their drive to complete, though this relaxed attitude may have been overdone at the first trial. In general, the students appeared to enjoy the experience, for there was no problem in encouraging them to speak, and some of the scenarios (notably *Compatible friends* and *4 characters*) provoked frequent laughter. Their responses to the questionnaires which they filled in at the end of the exercise were on the whole positive (see later report and *Table 6.1*).

The scenarios were written without difficulty in the time available and successfully fulfilled the criterion of local relevance. A complete set of the materials is given in Appendix 6.2: a brief account of each scenario now follows.

Introduction to Dorset

- situation: new students are arriving on Friday, activities to be suggested for them to do before they start the course on Monday.
- materials: colour pictures of heads, events list.
- goal: agree a timetable within the class for various activities.

A job in Dorset

- situation: student has been invited to stay on after the end of the course: he needs to find work to finance it.
- materials: advertisements from the local press.
- goal: discuss the jobs and who would be good at them.

Compatible friends

- situation: a friend from a past holiday wants to get in touch again: shall we meet or not? (Adaptation to this course book of an earlier scenario.)
- materials: 4 postcards for each group, sets of notes for phone calls.
- goal: decide whether or not to meet the friend, then tell him the decision on the phone.

Lecturer

- situation: a lecturer is to be invited by the students to give talk to the school.
- materials: outline information about 12 people, a biography of each.
- goal: agree who is to be invited to give the lecture.

Books for the library

- situation: several books have been donated for the school library and the students are to choose which would be most suitable. (Adaptation of an earlier simulation.)
- materials: 8 paperbacks, representing a wide variety of topics and styles.
- goal: agree as a class an order of preference.

Afterthoughts

- situation: students to write short stories for the school newsletter.
- materials: none.
- goal: agree which of the stories produced is to be sent in to the editor.

Four characters in search of a drama

- situation: students to prepare a short play for the school end of term show.
- materials: pictures of 16 sculptures.
- goal: class to agree which play is the better (from 2 groups).

The scenarios worked well in general. The linking of tasks to the students' present situation was successful: those relating to Dorset (*Introduction to Dorset* and *A job in Dorset*) could be seen to connect in students' minds with places and events they were already familiar with; a visiting lecturer *(Lecturer)* featured regularly in the weekly events list; and finally, without the prior knowledge of the researcher, the story for the school newsletter (*Afterthoughts*) occurred just as the actual school newsletter was published (with this class teacher as editor) and the short plays (*Four characters*) were prepared at the time when the school pantomime was in rehearsal.

One considerable success was the identification for transcription purposes of individual students as recorded on the tape. This had been foreseen as a potential problem, and early steps taken to resolve it were to allocate letters A, B, C to the groups and numbers to students within groups (A1, A2, A3, A4) on a systematic basis derived from square sets in dancing (where the couples are numbered 1 to 4 in an anticlockwise direction, with 1st couple backs to the band). In the same way, since there were never more than four students in a group, the student sitting back to the blackboard wall was always numbered 1, and the others identified in relation to him,

anticlockwise. The note sheets were coded in advance for each student with letter and number and handed out appropriately when the groups had settled. A plan was then drawn of the class positions with names taken from the note sheets as the students filled them in. In addition, the researcher made a list of names with their identification codes and wrote against each name a few words noticeably spoken by the relevant student so that the voice could be matched with an utterance recorded on the tape. A final check, if needed, was to match a student's writing on the note sheet with his speaking on the tape. This rather elaborate system eventually worked well: in the last three scenarios no intelligible utterance on the cassettes remained unidentified.

The greatest logistical problem however, as with the London trial, was in the recording of students' talk. The importance of this for the project was that methods of assessment were to be considered after the event, empirically derived from the data available, and for this procedure to be fruitful, enough student text (both spoken and written, but especially the former) had to be available for study. Better results could certainly have been obtained by recording the group discussions in a studio, but this would have been incompatible with one of the essential aims of the project, to set up a classroom-based system for assessment. In the event however enough spoken text was collected to provide data with which to illustrate various methods of assessing for diagnostic and achievement purposes over a variety of task types.

In this section the main empirical work for the project has been described: a series of trials which used the system of course book analysis and the rules for the design of scenarios to produce evidence about the application of the scheme in practice. This was done on a repetitive basis in one school, with variations depending on the content of the course books in use and the situation of the students in the classes. The results of this series of experiments are now to be discussed.

6.4 Outcome of trials

The data available from these trials were of three kinds: students' recorded spoken texts, their written answers on the note sheets, and their answers to the questionnaire. The most important of these is the first: what students say in response to the scenario is to be the main basis of their assessment, since the overall aim of the scheme has always been to work with interactions to provide evidence of student achievement. The written responses are less important, for they are in theory notes without an audience beyond the student who wrote them. The results of the third source of student production, the self-assessment questionnaire, were intended to achieve two ends: reflection by the student and information about the scenario for the researcher.

The spoken texts are considered in detail in Chapter 7, as input for the discussion of marking systems; accounts of the written work and the questionnaire results now follow.

For each scenario, three examples of responses to the note sheets are quoted in *Figure 6.1* These represent a somewhat random selection, but are intended to be typical of the written results overall.

As pointed out above, the notes are to be written by the student as reminders and prompts for himself, as instructed in the rubrics. However, they have a hidden audience (of which the students were no doubt well aware) in the person of the assessor, in this case the researcher. But the fact remains that writing was introduced firstly as a reminder for future talk for the student and only incidentally as a basis for an additional assessment. It was only in the course of the trials that the value of the written answers in locating and identifying individual students (by matching writing with speech) was appreciated and exploited.

What the students wrote is a clear expression of their own interests and of the relationships that they make with others, whether represented in pictures (*Introduction, 4 characters*), described in writing (*Lecturer, Compatible friends*), or in real life in the classroom (*Afterthoughts, 4 characters*). They respond positively to potentially real situations (*Job in Dorset, Library*), but can also involve themselves in affective and imaginative events (*Compatible friends, 4 characters*), seeing not only the implications of what is presented to them, but also showing skill in the language needed to explain how they respond (*Compatible friends, Library, 4 characters*). There are also indications that they are putting into practice effectively the language they have been learning (lexis specifically in *Introduction* and *4 characters* but also generally; structure in *Afterthoughts* and *Lecturer* specifically, but also again in general).

Figure 6.1: Writing from scenarios: examples

1. From *A job in Dorset*
I would like to look after baby. My favourities hobbies are go to the shopping, go to the beach swimming
2 animal assistants required
because I am fond of animals, specially I love cats. I happy because I need money and I have more patient.
I worry because I haven't experience, and the problem with the transport because I live far of Corfe Castle
Elvira has chose taxi driver because she has experience and she likes to drive.

2. From Books for the library
quite interesting, the writer tries to look at this world from a rabbit eyes
encouraging, enjoy it if read it be heard
no pictures, but not difficult
not difficult, easy to understand
the words are printed too small and quite long story makes me bored
quite famous, but I think it is boring
quite long, exciting, for English is good
not difficult, but quite a lot of vocabularies
Jonathan Livingston Seagull
younger, intermediate
I like the book which makes people to think
Watership Down
younger, intermediate
it is interesting and it isn't written down from human being's eyes
The River
younger, intermediate
I like to read a short story and easy to understand

3. From *Afterthoughts*
...that I told a rather fat friend of mine that she would survive the longest if there was a famine one day
I wouldn't say such a thing any more because it hurt my friend
Marlen told us she regretted not having said to another customer that it was her own turn
Daniel regretted not having checked the timetable before so that he was forced to take a taxi because there were no more trains

4. From *Four characters*
16
Jane
energetic, strong, mysterious, relaxed cold
beautiful, confident, independent
strength and independence

Self-assessment questionnaire

The request to students to comment on the scenario they had just completed was intended to achieve two ends: reflection by students on what they had done as an example of how much they could do with the language at their disposal; and at the same time information for the researcher on the success of the scenario as a mechanism for provoking language use, as reflected in the views of students attempting them. The self-assessment aspect of the questionnaire was also seen as a potential basis for subsequent class discussion, as part of the intended feedback to students, but was not used in this way in the trials.

An example of a questionnaire is given in *Figure* 6.2.

Figure 6.2 Self-assessment questionnaire: content (from *Introduction to Dorset*, as an example)

1. Class assessment	
I think that as a group we did	well/fairly well/rather badly
2. Self-assessment	
In the scenario,	
I could say what I wanted to say	often/sometimes/not often
I understood and talked	well/quite well/with difficulty
I found that talking about	
- people and relationships was	easy/a problem/difficult
- activities was	easy/a problem/difficult
- other things was	easy/a problem/difficult
- things that are happening now was	easy/a problem/difficult
- things that happened in the past was	easy/a problem/difficult
I found that discussing things with the other students was	easy/a problem/difficult
3 The scenario	
This is how successful I think the scenario was:	
- it was an enjoyable exercise	yes/not really/no
- it included a lot of what I learned during the course	yes/not really/no
- it showed me what I can do with the language I have learnt	yes/not really/no
4 Notes	

A summary of all answers to the seven questionnaires attached to the scenarios in the trials is given in *Table 6.1*. The form of the first three questions remained constant for all scenarios: they dealt with how well the learners did as a group and how well each individual felt he had performed. The next three questions were amalgamated under the heading 'lexis', and varied according to the demands of the specification for a particular scenario. For example, with *Introduction to Dorset* the topics were 'people & relationships' and 'activities', together with a catch-all 'other things', whereas with *Lecturer* the topics were 'people's personalities', 'arrangements and excuses', again with 'other things'. The two following questions dealt with tenses, explained without the use of metalanguage: 'things that are happening now' and 'things that happened in the past': these were collected together under the heading 'structure'. The next question, 'discussing things with the other students',

was constant in all versions, and headed 'discourse'. The last three questions were concerned with how effective the scenario was considered to be as a reflection of the individual's learning, and were again constant over all versions. In later questionnaires, a comparison was asked for with previous scenarios, since some learners took part in more than one. A few further variations with questions related to individual scenarios, for example one on speed reading for *Lecturer*, are omitted from the data in *Table 6.1*, as being unrepresentative. All are general questions, but they give a good impression of what the learners felt about their involvement and the success or otherwise of the scenario as a system.

In *Table 6.1* the total number of learner responses is given for each heading, varying according to the amount of amalgamation involved, and then four columns are listed, representing percentages of positive, uncertain, negative and nil responses respectively.

Table 6.1: Self-assessment questionnaires: data

question	topic	N	percentages of responses over all scenarios positive	neutral	negative	no response
1	as a group	69	61	39	0	0
2.1	could say what wanted	69	64	33	3	0
2.2	understood	69	26	64	7	3
2.3-2.5	use of lexis	220	70	24	4	2
2.6-2.7	use of structure	149	70	21	6	3
2.8	use of discourse	69	51	15	2	1
3.1	enjoyed scenario	69	65	32	3	0
3.2	reflected learning	69	26	61	12	1
3.3	showed can do	62	47	44	9	0

The first fact to be commented on is that the responses are largely positive, perhaps because of the happy atmosphere of the classes or even out of a desire to please. But there are some relatively high returns for the middle (neutral) category, notably for *understand* (64%) and *learning* (61%), and a more balanced distribution under *can do* (47/44/9/0), and this seems to indicate that the students were thinking about their responses, not giving general, blanket approval to themselves (and/or the researcher). One interesting point about the more negative responses is that they tend to deny the aims of the project, indicating that although students enjoyed the scenarios as activities, most did not see the exercises as using specifically the language they had been learning immediately beforehand, nor as particularly useful in showing them what they could do with what they had learnt. It may be that students do not recognise in retrospect when they have been using language they have internalised: detailed analysis of the spoken texts produced in response to the

demands of the scenarios might indicate how what they said related to what appeared in the analysis of the course books, but this is beyond the scope of the present account.

6.5 Summary

The development of scenarios worked well as a gradual progression from the experimental prototype, which used local materials to plan activities for a visitor, to the more demanding scenarios requiring an imaginative response. For example, the first two in the trial series deal with familiar topics such as entertainments and jobs, illustrated with pictures of people and advertisements, whereas later scenarios (particularly the last) require imaginative recreation of character and place and depend much more on the learner's individual outlook and experience. There was a parallel advance in the administrative procedures for the trials, including for example the timetabling of analysis and materials writing, the sequence of events in the scenario and identification of students in the classroom.

On the debit side, though various forms of check sheet were devised and experimented with over the eleven trials, the original aspiration to provide some form of reporting which would give immediate feedback to students on their performance was unfulfilled. The alternative procedure suggested, reporting back to learners by means of written statements representing what they have achieved in taking part in the preceding scenario (to be described in detail in the next chapter), is only one kind of feedback, and further developments may be envisaged such as using some adapted version of the marking system to enable contemporaneous assessments to be made by a colleague, or more traditionally, playing back to the learners extracts of video recordings of their contributions and commenting on them.

The scenarios themselves were effective in practice as a means of engaging students in a sequence of interactions. The essential 'motivational push' was nearly always evident, and students reported positively in the main on their experience. The format eventually stabilised as a single note sheet for participants, with a self-assessment/evaluation questionnaire on the back, and this worked well both as a record of what had been done by the students and as an indication of their reactions. The main achievements of the trials however were the translation of specifications into tasks and the matching of scenario content with the current context of the students. In addition, circumstances forced consideration of a abridged and therefore more economical version of the analysis system which was adapted for further trials as detailed in Chapter 7.

The questions originally asked of the trials in schools were:

1 Do scenarios produce responses from learners which are relevant to the previous course material?

2 If so, are the the responses sufficient for assessment purposes, both as to quantity (enough) and quality (differentiation)?

3 What form of mark scheme provides the most useful information for learners and teachers?

But the information which would answer these questions remains for the moment inaccessible, for until some interpretation of the responses of learners to the scenarios has been undertaken, they cannot be judged for their relevance to the previous course material or for their sufficiency in terms of assessment. This interpretation rests on the application of a marking system to the transcripts, which is the burden of the next chapter.

Chapter 7

Reporting back to learners and teachers

7.0 Introduction

In recent chapters, the concern has been materials which refer back to preceding work in class and at the same time lead forward to assessment. This chapter now investigates using the outcomes of these materials, realised in the form of scenarios, as information for feedback to learners and teachers. There are two main issues: to find principles on which appropriate assessment may be developed; and to set up a system which will give useful information about participants' performance. Various approaches are discussed, such as adapting the course book analysis to provide categories for assessment of spoken texts produced by learners, and studying the texts themselves for indications of suitable assessment criteria. In this case, the main problem is how to interpret the texts so as to offer useful information to learners and teachers on the learners' performance.

It was proposed earlier that no detailed decisions should be made about marking systems until a series of scenarios had been trialled in the classroom to provide evidence about how learners responded. The only rule laid down in advance was that there were to be no rules governing the content of learner responses: participants were to use whatever language they had available to fulfil the demands of the tasks set. The evidence resulting from the use of scenarios is now available, in several forms: learners' written notes on the individual note sheets; their responses to the assessment questionnaires; comments from the researcher and teacher as included in reports; and recordings of the learners' spoken interchanges, with transcripts. As the commentary on the trials has shown, learners wrote a certain amount in the scenarios, but this was intended mainly to allow administrative checks and to guide the learners themselves in their group and class discussions later in the scenario. These written responses and the answers to the 'self-assessment' questionnaire were commented on in Chapter 6, and though there is some potential for investigating them as output, no further consideration is to be given to them here as potential sources for assessment, since the main preoccupation of the project has consistently been the assessment of spoken interaction between learners.

The following discussion is therefore essentially concerned with the assessment of speaking, but in a generous sense, to include the influence of others' contributions to a discussion and decision-making for a given purpose. As explained in Chapter 6, the scenarios were designed to include a range of different kinds of interaction,

from argument to narrative. At the same time, they were organised so that learners talked in a variety of social contexts, usually in a sequence from discussion in pairs (for clarification of the situation and their aims within it), to a group discussion and then to coordination of their ideas with the class as a whole. The different kinds of talk resulting from this sequence, and the range of tasks set by the scenarios, together produced a variety of texts which will now be sampled, to guide a discussion of marking systems. The intention is to set up principles on which judgements are to be made about learners' production and to devise a mark scheme which will guide assessors in making relevant and consistent decisions.

The original intention was that there would be three parts to the assessment: interactions, achievements and group working. The first part was to report on what participants said in scenarios in order to fulfil their aims, with results in the form of statements, for example *You understood what was expected* or *You started most of the discussions*. The second part was to be a record of what means participants used to put their aims into effect, with results in the form of a diagnostic report on language, for example scores for successful and/or inappropriate use of lexis and structure, to be interpreted by the teacher in terms of accolade or remedial suggestions. The third part was to provide further statements, this time on each participant's contribution to the group, for example *You had useful things to say but you need to listen more attentively to what your partners are saying*. As always in this project however, these elements evolved beyond the original concepts as work on the mark scheme progressed. An account of these developments now follows.

7.1 Marking systems

Some historical approaches

Pollitt (1991) identifies two assessment strategies, counting and judging, and provides examples from different approaches to assessment in sports. Counting is the method used in events like the high jump, in which competitors attempt higher and higher tasks until they fail; judging is the kind of assessment used in ice skating, where experienced observers make spot decisions on the basis of a long apprenticeship towards a consensus of experts. In the assessment of language, Pollitt argues, counting is based on adding up scores (item-based testing) and is mainly concerned with the difficulty of questions, making the assumption that all acceptable performances on any particular item are equally good and setting as the criterion the *quantity* of acceptable performances. Judging, on the other hand, is based on the rating of a performance, with a concern for the precision of rating criteria, making the assumption that all tasks are equally difficult and setting as the criterion the *quality* of performance (Pollitt 1991:52). The present project, starting from the dual aspiration of providing both diagnostic and achievement feedback, now needs to find adequate means of fulfilling these aims, with the transcripts as evidence to substantiate any proposals. Pollitt's counting and judging strategies seem at first sight to fit these needs rather accurately: diagnostic feedback may be

based on the counting of details in learners' output, whereas achievement may be reported in terms of judgements on successful interactions.

Counting is applied in the more 'objective' types of test, where a score represents an accumulation of correct answers. But the questions to be answered have already been subjected to *judgement* as to what should and should not be included as content, and often also as to *value* on the evidence of pretesting. This is the point of Pilliner's (1968) observation that 'objective' tests are objective only in the marking, for considerable subjectivity has gone into their preparation. Procedures of this kind are not readily applicable to the assessment of spoken language, where even though an interview may consist of prescribed questions, the answers are by their nature largely unpredictable, and judgement is required in deciding what value to attach to any given answer.

In practice, a judging approach is appropriate for oral testing because it can be used for relating performance to a set of criteria which are allotted marks by a mark scheme. The award of these marks is within the examiner's gift and in awarding them he is exercising judgement. Historically, as in the FSI oral interview (Jones 1979), the total score is an accumulation of judgements on such criteria as pronunciation, stress and intonation, grammar, vocabulary and fluency (these are commonly quoted in books on testing for teachers, eg Heaton 1975:94). But within this 'category approach' there may be an attempt to define points at issue which have been isolated as objectives in a specification and so can be placed before the candidate to see if he can deal with them. For example an oral examiner may work through a series of tenses, or attempt to 'force' particular structures by the nature of the input, or use pictures to test specified vocabulary. To this extent, the assessment may claim to be systematic and to represent a stated sample of content. Page (1983) offers a thoughtful discussion on the advantages and disadvantages of a 'defined syllabus', suggesting somewhat reluctantly that lists are necessary for external examinations in schools. The present case is different because it rests on the notion that assessment should be organised to allow the learner to show what he can do with whatever resources he can muster.

More communicative approaches demand more integrative marking so that there is less demand for specific responses. Marking criteria have been developed instead as headings for elements arranged in scales, for example the 'accuracy, appropriacy and fluency' of the RSA model (Morrow 1979) which in turn became the 'degrees of skill' of the CCSE of 1990 onwards (see Chapter 2). More recently, criteria for assessment have been reinterpreted as 'facets' of proficiency (Bachman 1990), leading to scales for scoring which are 'based on theoretical definitions of the construct' and 'referenced to specified levels in different areas of language ability' (Bachman & Palmer 1996:211). Developing theory into practice, UCLES has designed a current model of communicative language ability on the basis of the Council of Europe definitions for Threshold Level and Waystage (van Ek & Trim 1990) as well as the work of Bachman (1990) and Canale & Swain (1980) (Saville &

Hargreaves 1999:46). Marking is done according to rating scales with defined criteria, derived from the UCLES model (itself derived from Bachman's (1990) work): grammatical, discourse and pragmatic competences and strategic competence (Saville & Hargreaves 1999:45).

Present intentions

The systems discussed above all depend on judgements because they are concerned with the rating of performances, and the present case is no different from them in its concern with the assessment of speaking. They all claim to assess candidates' proficiency, that is, their ability to operate in real situations. But the intention of the present project, to provide three different kinds of information - on aims expressed, mechanics of language use and group working - implies different approaches to marking. In Chapter 3, after a review of various marking systems for spoken language, it was suggested that there were two appropriate systems for present purposes: those based on 'units' and those resulting in 'statements'. The unit system is essentially a means of attaching a numerical value to each event or utterance within a sequence of talk, a variation on the traditional mark scheme which ends in a score, but in the case of units, a positive award for communicative value on a rising 3-point scale rather than a score judged out of a maximum of (say) 5 points on (say) 5 criteria. The unit system can be an economical means of providing diagnostic information about learners' use of the mechanics of language, as for example in the Ilyin Oral Interview (Ilyin 1976), which is based on a picture sequence and is scored on a 0-2 scale, with one point for appropriateness of response and one point for structural and grammatical accuracy. The T level test (Groot & Harrison 1979) marked each candidate answer, or question, on a 0/1/2 basis in which the first mark was awarded for a successful communication and the second for accuracy.

The second suggested system, resulting in statements, is a means of describing in words what has been done as evidence of learners' skill in using whatever language is available to them to cope with situations. This is essentially a judging procedure. An early example of this kind of rating is the SKF Stages of Attainment Scale (ELTDU/OUP 1976), which provides information about the language ability of personnel in industry, presented in two sections to meet the differing interests of two kinds of reader: administrator and teacher. Each section consists of 'Can...' statements at eight levels in three skills areas (Listening & Speaking, Reading, and Writing). Another example of this kind of procedure is the matching of student achievement with the IBM scale (reported in Trim 1978:68), for example 'Can write letters making arrangements for business appointments', 'Can function in a small group planning projects' (these are two of some 100 statements at various levels of achievement). Similar judgements are made in using scale statements (North 1995), some of which may be empirically derived (Chaloub-Deville 1995, Fulcher 1996).

In the present case, the diagnostic assessment is intended to be a detailed report

on what the learner does with language to express himself and how accurately he does so. The achievement reporting is to be more broadly based, because it involves interpreting what the learner says in terms of the meanings he needs to convey. In both cases, the marking system is to be derived from samples of learner text, with the primary aim of finding out and judging *what* the learners have done, not *how well* they have done it. This turns the original counting/judging distinction inside out, for judging is normally concerned with how well and counting with what, but the reversal is necessary if learners are to be given factual information on which to act - what they succeeded in and where they fell short - rather than ratings which provide no particular instances. The concept *how well* implies a comparison with others and a relationship with some external standard. It is also likely to engender statements which include unspecifiable modifiers such as *fluently, accurately, effectively; simple, routine; less, more; and some, most,* which are not directly helpful in describing what has actually been achieved. In this spirit of directly relating assessment to performance, statements will report what a participant *has done,* without the predictions implicit in statements which claim to report what a participant *can do.*

The outcome of this discussion is a confirmation of the proposal that the reports back to learners from their participation in a scenario should be concerned with the facts of their performance. The most useful information for learners is likely to be an account of the language they have used in a scenario, followed by statements which give them indications of how they have participated with others, so that they see what strengths and weaknesses they have shown in their use of language for purposes which are relevant and practical for them as individuals.

7.2 Marking in principle

In order to arrive at a marking scheme, some principles need to be established on which the marking may be founded. As always, the starting point is what exists, which in this instance is learner talk as it is recorded. In addition there is, for research purposes, the luxury of transcripts, which are not usually feasible for the everyday. The data transcribed from the trial series, as described in the previous chapter, consisted of seven extracts (labelled A to G) from the recordings made of participants' interactions in various scenarios. This collection of material was the evidence called on for an extended investigation of marking possibilities, including a viability exercise with expert colleagues.

The first problem in the development of a marking system is to find ways of setting up practical principles which, when applied to learners' output, will provide useful information for them and for their teachers. As before, discussion of salient topics took place in working papers which explored alternative approaches and solutions, and the following discussions consists of extracts from these papers, suitably edited.

Tallying events

An essential element in the assessment of speaking by a teacher in a classroom setting is the time required to carry it out. Spoken language could be metaphorically represented as the continuous production of a ribbon of events involving both expression and comprehension, even if woven by more than two participants, all of them working together contemporaneously and maintaining contact in real time. A few lines of a transcript from one of the scenarios, as given in *Figure 7.1*, will serve as an example of what the marking needs to cope with. At this point, the transcripts were simply the result of listening to the recorded cassettes and copying down literally as much as could be clearly understood, adding line numbers and clarifications [in square brackets] and including the occasional undecipherable phrase (<u>underlined</u>).

Figure 7.1: Sample of transcript

line	*ref*	*text*
01	A1:	saloon stock cars I think she [pic 14] will not be very interested in it
02		because she's more a
03	A2:	yeah
04	A1:	a one to look nice
05	A2:	I think this one perhaps - they go away to the <u>height</u> Funskating attraction
06		[reading events list] or er or even the windsurfing it can be...
07	A1:	to windsurf
08	A2:	yeh yeh
09	A1:	and from Italy she like the sea
10	A2:	sure
11	A1:	so it's done

For someone listening and responding to a partner, as in this example, the process of understanding includes a particular kind of assessment of what is being said, involving judgements about innumerable aspects of the speaker's intentions (of which only some may be appreciated by the hearer) and about his means of expression (which may be interpreted in different ways by different hearers). In spite of this complexity, understanding nevertheless requires an immediate response, even if this is not verbalised. In the sample above, A1 appears to be making the running on the whole, with A2 mainly agreeing (lines 03, 08 and 10) but in the event it is A2 who offers a suggestion (05-06) which is elaborated by A1 (09) and then adopted as the answer to the task (11). The partners have clearly understood each other, processed each other's successive contributions, made immediate judgements about appropriate responses and then spoken those responses within the short time allowed by continuous conversation.

Assessment by an external observer in order to construct feedback information for learners is similar in principle but different in kind, for the response required demands a different kind of judgement. But one approach to a marking system for a live assessment could be to suggest that the same speed of interpretation can be learnt by an observer as by a participant, even if the use made of the resulting understanding is different. If meanings can be exchanged at a given rate, including hearing, interpreting the sounds heard, extracting the content, then formulating a response mentally and externalising it in speech, might it be possible for a similar rate of transfer to some scoring or note-making system to be achieved by a listening but non-participating assessor? With practice, might categories of judgement be applied within the time scale of an unhurried conversation?

The implications of assessing contemporaneously with learners' speaking in the classroom were explored in another working paper, which began with a review of external observation procedures which showed that systems set up to analyse classroom behaviour in action are usually too closely related to pedagogic interests for present purposes. They tend to be concerned with such topics as how the teacher controls class discussion and how students respond to the teacher's moves. In observation systems of this kind, Fanselow maintains, learner initiative is to be expected only 'in a classroom setting without discipline' where 'students also react and solicit' (Fanselow, in Allwright 1988:132). 'In fact, one central purpose of all of this systematic description is to begin to identify consistent relationships between characteristics of communication so that we can begin to base our teaching on evidence of effectiveness plus theory rather than on theory alone or one's whims' (op cit p 142). Fanselow's preference for a systematic structure is impeccable, but his concern here is teacher rather than learner effectiveness.

Some means needs to be found of representing learners' rather than teachers' interests in the application of an assessment system. Perhaps categorising learners' output in some way would provide a basis for the necessarily rapid decisions required for contemporaneous judgements. Flanders, for example, envisaged that observers of classroom events would memorise all 10 of his categories of teacher-pupil verbal interaction and tabulate a unit of behavior approximately every three seconds (quoted by Allwright 1988:37). A timed spot check system of this kind, as extensively used in classroom observation (for examples, see Croll 1986 and Allwright 1988), might be adapted for the assessment of language in use.

A scheme for analysing events in language classes was proposed by Allen, Frohlich & Spada. Their COLT (Communicative Orientation of Language Teaching) observation scheme is in two parts: Part A describes classroom events at the level of activity and Part B analyses the communicative features of verbal exchanges between teachers and students as they occur within each activity (Spada & Frohlich 1995). Both these sets of observations take the organisation of the classroom as a starting point. But it is not accidental that the scheme's title refers to language teaching (rather than to language learning): again the focus is on the classroom as an

instructional environment, whereas the interest of the present project is in the performance of each learner as a language user rather than as a class participant, even when he is interacting with another learner. Another difference is that in the present case the structure of the exchange is regarded as set up by the demands of the course book rather than by the intentions of the teacher, which may seem similar, but are by no means identical. The COLT check lists propose slots to be marked rather than comments to be made, so that there is room for as many as 80 categories to be considered for every event observed in Part B. This is too crowded a menu for present purposes, where in the main judgement is to take precedence over counting. Nevertheless, the COLT scheme has many coincident points with any effort to represent communicative events in the classroom and sets some helpful pointers for directions in which the development of assessment procedures might run, mainly in terms of categories for consideration.

Another potentially helpful example of classroom observation techniques, the *One in five* study (Croll 1986:18ff) suggested further ideas. This study involved the recording on a checklist of individual children's classroom activities and interactions at 10-second intervals. Coding was carried out for 33 categories under 6 headings, covering teaching organisation, reading, curriculum content, child activity, pupil interaction and mobility. These categories were 'defined by the research questions which the observation was designed to answer', and resulted in scores representing the designated activities and interactions. In the present case, the hope was that definitions could be developed from the observation of learners in action rather than from a preconceived framework.

In an attempt to explore further the ways in which an assessment structure might be based on learners' production, various forms of tally sheet (an example is given in *Figure 7.2*) were devised for experimental use with early scenarios to capture instant impressions of some of the interchanges achieved by participants. The point of these was to explore ways of encapsulating what learners said in a discussion, treating it as a succession of contributions and at the same time making judgements on the success of each element in terms of communication. In this way, starting points for marking might be identified, leading to a categorisation of elements for which credit could be given.

Figure 7.2: Tallysheet

Tallysheet

mark 0/1/2: 0 = attempt, 1 = adequate, 2 = apt

student	events ⇨					total
1						
2						
3						
4						
5						

The interchanges were to be divided into 'events' and the three possible judgements on responses were to be labelled: '0 = attempt, 1 = adequate, 2 = apt'. Two main problems appeared: that the identification of 'events' depended on a wider study of learners' speaking output than was available from a single scenario; and that entering a number on a checklist to represent a category (as in the *One in five* study) is a snap decision which can only record an occurrence and not the judgement of a value, as is demanded by an 0/1/2 marking system. The tallying procedure was therefore overtaken by more detailed consideration of what the three awards (attempt, adequate, apt) might be made for, as reported below.

The central problem for any form of judgement, whether concurrent or retrospective, is to identify what are labelled 'events' in *Figure* 7.2 and to divide the text up accordingly. Functions might seem an obvious point of departure, but there can be no systematic definition of where they begin or end, and in any case there is always more than one function in operation at any given moment: 'Every adult linguistic act, with a few broadly specifiable exceptions, is serving more than one function at once' (Halliday 1973:34). The tallying system, though promisingly contemporaneous with the progress of the scenario, seemed impossible to put into effect unless workable definitions of the concept 'event' be could be found.

Such a definition might be derived from a system that already existed: the categories set up for course book analysis, as described in detail in Chapter 4. The question is whether the definitions arrived at there might somehow be adapted for marking purposes, so linking directly back to earlier principles and at the same time following the logic underlying the analysis through into assessment. A detailed exploration of this possibility was made, and though some positive results were identified, there were considerable drawbacks, including doubt about reliability and the number of relevant statements feasibly to be derived from the procedure. A further difficulty was that an adequate range of achievement could not be covered by adaptations of this kind. But this exploration of AS1 and AS2 as sources of assessment criteria did suggest directions in which further development could move. It showed that it could be feasible to derive statements of learner achievement as logical interpretations of what they said, and that levels might be considered as an issue of comparisons within the data rather than a relationship with an external structure imposed from outside.

Learner aims as a source of statements

The next move was to make more direct use of the language that the learners produced in response to scenarios. The chief argument supporting this idea, apart from its link into the work of the project so far, was that it recognised the learners as the originators of the transcript and hence the primary source of the statements. Seven extracts of cogent learner texts, taken from the 600 lines of transcript and labelled A to G, were used to provide a varied basis for a discussion of marking systems.

The transcripts were first considered as records of learners' aims, such as *identification* (of somebody or something), *narrative* (a sequence of events) or *reasons* (for events or attitudes). The next step was to set up some system for dividing up the text to represent aims that could be represented in statements. One approach could be to set arbitrary boundaries, for the principle of arbitrariness helps to promote consistency. It has precedents in the world of sport, for example, where rules are established to ensure fairness rather than logic as a basis for unbiased judgement (Pollitt 1991). In the present case, bias could occur if some external system of division was applied which was related to teacher-oriented logic, which might allow classroom business to take precedence over learner production. It was therefore decided that the learner texts should be divided into lines on an arbitrary basis, using as a horizontal boundary the measure (width of line) of the word-processing program in current use.

In practice, however, this approach represented learner aims at too low a level of generality: a larger chunk of language was required as a base unit, and the use of width of line as a basis for text division could not operate as a substitute for rational linguistic groupings. What this exercise did achieve however, was an indication of what more needed to be included than the traditional concept of 'function', for example interactions such as self-correction, assistance from other participants, changes of approach and problems with formulation. It was also an important step forward in establishing what learners say as the starting point for setting up a marking system, in line with suggestions made by Fulcher (1996) and others, that assessment should start from the language data rather than be imposed on them.

Units for analysing spoken text

If arbitrary division of learner texts is not a viable solution for assessment purposes, some other system needs to be devised of doing so rationally, generating manageable units, or chunks, within which assessments may be made of each participant's contributions to a scenario. Such a system could start from consideration of either speakers or text. Speakers take turns, or pause within a narrative or argument either to elicit acknowledgement or to take a breather (physically or mentally): any of these breaks in the text could be used to designate boundaries. Alternatively, since a text is structured by grammar or discourse, or both at the same time, these textual shapings could be used as guidelines for the grouping of words. Analysis of spoken text, whether for discourse analysis or any other linguistic purpose, has always required some system of division into units on which comment is to be made. Perhaps some of the same principles could be applied in the development of assessment procedures for learner texts. The ideal unit for dividing up learner texts would be one which provides a chunk within which some relationship can be discerned between the actual words (or, in the extreme case, sounds) which are used and the purpose of the speaker, even if this purpose can only be arrived at by logical inference. For example, Chaudron (1988:45) defines

a *T-unit* as 'any syntactic main clause and its associated subordinate clauses' and a *communication unit* as 'an independent grammatical predication; the same as a T-unit except that in oral language, elliptical answers to questions also constitute complete predications'. Some further rules are needed to guide decisions for the present project and hence to improve consistency in the division of text into units, specifically what forms indicate the 'independence' proposed by Chaudron.

Brock (1986) defines a communication unit - which she calls a *C-unit* - as a basis for investigation into the length and complexity of learner answers to 'open' as against 'display' questions, the former being questions which are genuine enquiries to which the asker does not know the answer and the latter those which allow the responder to demonstrate skill with language rather than supply useful information. The application of C-unit chunking to learner discourse is usually in the context of research into the nature of learner talk, for example the language learners use in response to different types of task (eg Foster & Skehan 1996, Skehan & Foster 1997), rather than as a means towards articulating assessment. This is probably because the C-unit seems at first sight too small a chunk for assessment purposes, either retrospectively or, especially, for instant judgements. But if it could be used experimentally as a starting point for the analysis of transcripts it might accumulate around it practical pointers towards a chunking system which could be applied in the context of assessment, and might even suggest ways in which it could cohere into larger and more manageable units. This was the assumption driving the next attempt at transcript interpretation.

As before, principles and then practice were argued through on paper: in this case a methodical justification for using C-units in an assessment system derived from the learner language which was available in the transcripts. The explorations of the transcripts showed that neither Chaudron's definitions of T- and C-units (as above, 'an independent grammatical predication' centring on grammaticality) nor Brock's (1986) C-units, as used by Foster (1993:8) ('utterances which are meaningful but not necessarily syntactically complete') are adequate for present purposes because they do not offer enough guidance on how learner aims are to be allocated to linguistic events. The notions of 'independent' and 'meaningful' are problematical, because both need to be considered in balance with context. What is independent in a grammatical sense may be part of a larger meaning (for example *and we each know each other from several er letters so we know quite a lot of each other* [Extract B] could be considered as two independent grammatical clauses, but they are evidently part of the same mental concept, or 'meaning'). On the other hand, what is meaningful may contain no syntax at all (for example laughter conveys understanding and appreciation and therefore should qualify somehow for inclusion under 'meaning', even though it has no language content).

The extended definition of *Figure 7.3* is therefore to be applied in marking transcripts.

Figure 7.3: Definition of a C-unit

> As applied to the learner texts of the present project, a C-unit may consist of:
>
> - an independent grammatical predication consisting of a main clause and its associated subordinate clauses; or
> - a grammatically incomplete but elliptical statement or response which conveys a recognisable learner aim; or
> - an utterance which conveys a recognisable learner aim whether or not in the form of words.

The value of the third condition above is in making room for the assessment of responses furthering the interchange with sounds such as 'yeah', 'mm', corrective murmuring within the group, and laughter. These express understanding, encouragement and other supportive moves and need to be credited to the individual or group, as appropriate.

Aims as a basis for categories

With this new definition of the C-unit as the means of dividing up the transcripts into assessable chunks for present purposes, it is time to return to the discussion of what these chunks would contain. The starting point was originally to take the learner's aim as the most important consideration in applying assessment to transcript, since if it could be identified it would centre the assessment on the individual.

The development of an assessment system on the basis of allocating aims to chunks went through several stages, elaborated in a further working paper. The first step was to search through the transcripts in an attempt to identify and list all participants' aims. This first list, drawn up in order of their appearance in the transcripts, resulted in a random and unmanageable collection. But it soon appeared that two more general varieties of aim could be distinguished: factual and affective. Under the heading 'factual' appeared 19 nouns, for example: *acknowledgement, agreement, clarification*...; under 'affective', 16 nouns, for example: *acceptance, accusation, apology*.... Eventually there was found to be a need for a third category of events which were not part of the learner's communicative aims as such, but represented external factors in the structure of the interchange, such as information needed for the interpretation of the text (for example references to pictures or realia, prompts from one participant to another) and reference to administrative concerns. These could be listed as either nouns or adjectives: *commentary, [incomplete], prompted, self-correction*.... - 17 of these events appeared.

At this point the list was converted into a more coherent matrix in which headings were allocated to four categories, with events relating to areas beyond the task subdivided into those concerned with the discussion and those resulting from the tape-recording process. These four categories were: *factual, affective, interaction, ex-task*. Each of these categories was considered as deriving from a different factor in

the interchange: situation, learner, discussion and recording respectively. In an attempt to clarify the suggested categorisations, the contexts in which the transcripts were generated were defined as scenario, individual, social and physical. Although this discussion helped to define the factors found in transcripts, the main problem with the matrix was that it contained too many elements if the system to be efficient, that is, easy for for assessors to comprehend and apply consistently .

In an effort to redefine the relationship between the headings and texts therefore, and at the same time to reduce the number of elements and categories, a further revision was undertaken. In the course of it, the question arose as to which among the elements under discussion were likely to be specific to a scenario and which might have more universal application, for this could affect the appropriateness of different assessment methods. Linked with this discussion were further thoughts about differentiating the assessment of the individual from the assessment of group working. As a result of these ideas, a distinction was made between interactions which were concerned directly with the task (eventually to be known as *transactions*) and those which reflected participants' cooperative moves in completing it (*activities*).

Summary of principles

The point had now been reached at which the practicalities of marking could be addressed. It had been argued that learner texts were to be divided up into units (in this case, the modified form of C-unit as defined in *Figure 7.3*); each unit was to be labelled with an interaction encapsulating the intention of the speaker, such as *agreement* or *clarification*; each scenario was to be allocated further interactions relating to cooperation between participants in completing it, such as *resolving conflict* or *encouraging*; and a diagnostic element was to be included for each participant, based on the language he used to put his aims into effect, referring specifically to structure and lexis. These decisions fulfilled the proposals made at the beginning of this chapter (though set out in a different order) that that there should be three parts to the assessment, all concerned with various kinds of participant engagement - with task, with language and with other participants. These were now to be combined in a workable marking scheme.

7.3 Marking in practice

Though the assessments of *transactions* and *activities* were both intended to result in statements, they needed to be arrived at in different ways. The main distinction between them was that the first referred directly to the learner's aims in dealing with the task in hand and the second referred to how he cooperated with others in reaching solutions. The statements referring to aims would by their nature apply to the learner's performance in any scenario, for they would record what transactions he had carried out with his available language - a sample of his total capacity at this point. The other kind of statement would be specific to the current

scenario, since its demands could require such differing cooperative inputs as 'keeping the discussion going' or 'responding quickly', depending on the nature of the task set. (The third kind of assessment, of the learner's use of language, is to be considered later, since it is concerned with diagnostic interests - means rather than ends.)

There was a related distinction to be made in the way each of the two kinds of statement was arrived at. The transactions were to be listed as a bank of learner aims from which the assessor was to draw an appropriate label for each unit of text, whereas the list of activities was restricted to those allocated to the collaborative work appropriate for the particular scenario. For activities, an assessment was to be made on the basis of a scan of each participant's transactions, then transferred directly into statements.

Universal and particular interactions

The result of all this ratiocination is two lists, as given in *Figure 7.4. List a*, now known as *universal* interactions, are learner aims derived from inspection of all the transcript extracts, as explained above, but sifted through and amalgamated into fewer elements where possible. They are arguably common to all scenarios, if a wide enough range of activities can be considered to have been sampled by the scenarios devised for the trial series. The interactions are restricted by two considerations: that they reflect aims (*what*) rather than reasons (*why*) and that they include only those which are susceptible to conversion into statements. To illustrate this requirement, *Figure 7.4* includes an annotation against each interaction which offers a possible starting point for the writing of a relevant statement. *List b*, to be known as *particular* interactions, are the demands made of participants by the pair or group work of a given scenario, and are not necessarily applicable to other scenarios - though many of them may be, depending on the nature of the discussion involved.

Figure 7.4: Universal and particular interactions

List a: Universal

- initiation	*starting off a discussion, a new topic*
- acceptance	*understanding and working with others' output*
. implication	*making or taking a meaning beyond the literal*
. empathy	*showing understanding of other participants' feelings*
. interpolation	*oiling the cogs of discussion*
- advance	*making a new contribution to the discussion*
. eloquence	*using words to good effect*
. emotion	*expressing personal feelings*
. enquiry	*asking for information, for others' views*
. offer	*helping the discussion on with information, with opinion*
- response	*taking up the discussion in answer to another participant's input*
. denial	*disagreeing with facts or views given by other participants*
. substitution	*replacing own word or phrase with another*
. first aid	*asking for or giving missing language*
- conclusion	*finishing the discussion*

List b: Particular

describing
suggesting
keeping it short
encouraging
identifying
characterising
questioning
finding reasons
resolving conflict
responding quickly
understanding
keeping it going
provoking response
linking
originating

The first intention was that the *universal* interactions would be available as a set, one interaction being allocated by the assessor to each C-unit in a transcript. These sets would consist of extracts from an overall list, each set varying according to the demands of the scenario but overlapping with other sets as necessary, so that there would be in effect a small bank (drawn from a larger comprehensive bank) of permitted interactions for the assessor to allocate for each scenario. This would reduce the task of the assessor by restricting the range of possibilities open to him. With the same aim in view, and also because less variation was possible in the demands of group work, four *particular* interactions were to be allocated to each scenario in advance to represent the performance expected of the pair/group as a co-operating unit. To help the assessor differentiate between universal and particular interactions, the former are listed in the mark scheme as nouns *(initiation, acceptance, advance...)* and the latter as participles/gerunds *(describing, suggesting, encouraging...)*.

Both kinds of interactions were allocated experimentally to each C-unit of the eight extracts from the transcripts so as to explore how successfully they could be applied to texts. This exercise showed that there was considerable overlap in the sets of *universal* interactions, and separate lists for each scenario were therefore abandoned in favour of one consolidated list, in the interests of simplicity. In addition, a single list, if it could be successfully applied, would promote a claim to generalisation. The single list is therefore used in the draft mark scheme (as exemplified later on, in *Figure 7.10*). The *particular* interactions remained in a separate list, from which four were to be selected for each scenario in advance as before.

Diagnostic assessment

The discussion reported in recent paragraphs has been working towards statements which would represent what participants in scenarios had done in terms of task fulfilment. But diagnostic assessments were also to be made on the same material, to provide learners with feedback on how accurately they had used the language at their disposal in participating in the scenario. The next requirement was therefore a mark scheme for this purpose which could be applied in parallel with the assessment of interactions.

The starting point for recording the diagnostic element was the 0/1/2 method (illustrated in *Figure* 7.2 by the labels 'attempt/adequate/apt'). However, this blunt instrument, though useful and appropriate for general judgements on utterances, was not explicit enough for reporting on what participants had actually done in terms of language so that they could be given useful feedback. To meet this shortfall, it was decided to record success in the use of language under two overall award categories, *weaknesses* and *strengths*, and then subdivide each of these categories into *structure*, *lexis* and *move*. This gave scope for credit to be given (or denied) for three aspects of each learner utterance, C-unit by C-unit. The units would already have been allocated an interaction from the list, and detailed accounts of the learner's use of structure, lexis and move within the same interaction would indicate what had been done in order to carry out the interaction. Successful *interactions* could then be translated into statements while strengths and weaknesses in the use of structure and lexis could be reported from scores for *achievements*.

The next consideration was how to generate these scores. It was decided that the scoring for lexis, structure and move should be done with ticks (✓) and crosses (x) on a 5-point scale (✓✓ ✓ 0 x xx) in preference to the simpler 0/1/2 basis, so as to improve discrimination by providing for better and poorer (rather than merely adequate or apt) performance. This was a compromise with the principle that marking should be positive, working up from a zero base, but was necessary if inadequacies were to be indicated in enough detail for recommendations to be made on remedial work. The idea of using 0 as base and scoring down from it as well as up was derived from the English Speaking Board's procedures (Burniston 1982:28 - see comments in Chapter 2), which rules that examiners should start

from an assumption of competence - scored in the present case 0 - and add or subtract credit according to performance - in the present case moving to ✓ or x. The use of symbols rather than numbers was intended to combat the almost instinctive tendency for markers to translate numbers into total scores, thus moving into *how well* territory rather than remaining within the area of *what*. Consistently with this view, totals derived from occurrences within a performance (eg ✓'s for variety in the use of lexis or x's for repeated errors in the use of particular structures) were not to reported as scores to learners, but used by the teacher as a basis for review, drawing the learner's attention to notable factors in his performance. Diagnostic information would thus be included in the writing of statements, with remedial suggestions where possible. This development moved away from the simple application of 0/1/2 as a counting procedure into a system which could provide for finer judging and at the same time be integrated with the judgements on interactions, since both *interactions* and *achievements* could now be reported in statements.

Contributions to group working

The third part of the assessment was a judgement on the effectiveness of individuals in contributing to the pair or group work involved in each scenario. They are the transactions given in *Figure 7.4* as *List b*, the bank of demands on participants for a particular scenario, representing the various *activities* required. The allocation of activities to each scenario depended on the kind of discussion demanded, for example the pair work in *Books* expected participants to understand rapidly what a book was about and agree on priorities, whereas *Afterthoughts* was concerned with narrative skills. The marking was to be made on the basis of a scan of each participant's interactions as noted for C-units, and allocation of the 5-point category system (✓✓ ✓ 0 x xx) to each of the activities listed.

7.4 Setting up a mark scheme

The next stage was to convert all this thinking into a practical mark scheme which could be applied by assessors on an experimental basis, to explore its viability. The marking was to be done in three parts, firstly allocating one of the list of relevant 'universal' interactions to a C-unit, then scoring its structural and lexical content on the five-point scale, and finally giving an overall score to the group on the four activities which had been designated as 'particular' for the scenario.

The following demonstration of the marking process is necessarily based on a transcript, though it had been envisaged from the start of the project that the end product would be a marking system which a teacher could use in the classroom, applying it to a recording or possibly even directly to the participants' contributions as they were made. This transcription-based demonstration is ideally only a first step in the intended direction, but will provide examples of how the mark scheme operates to put into practice the principles which have been discussed above.

A viability exercise was undertaken with a group of the researcher's colleagues. They agreed to apply the mark scheme to seven of the eight learner texts extracted from the transcripts of the trial series (the eighth was abandoned as too insubstantial for this purpose) and to give expert opinion on the scheme's working in practice. A set of instructions was written to explain what the interaction headings represent and what to enter in the C-unit boxes; how to score for structure and lexis; and finally how to assess the contribution of individuals to the discussion. The resulting comments are discussed later in this chapter; for the moment, the scheme will be explained step by step as an assessor would use it (and as the group of experts met it in the viability exercise).

Figure 7.1 gave a few lines of transcript from Extract A. These are repeated here for convenience as *Figure 7.5*. The text is taken from the scenario *Introduction to Dorset*. The participants, having looked at pictures of people who are going to come to the school as students, are now deciding (task 2) which of several activities (suggested in advertisements) might interest these newcomers. The participants are coded in the transcript as A1 and A2, as the first two members of group A.

Figure 7.5: Transcript: part of Extract A

A1:	saloon stock cars I think she [pic 14] will not be very interested in it because she's more a
A2:	yeah
A1:	a one to look nice
A2:	I think this one perhaps - they go away to the height Funskating attraction [reading events list] or er or even the windsurfing it can be...
A1:	to windsurf
A2:	yeh yeh
A1:	and from Italy she like the sea
A2:	sure
A1:	so it's done

The first task for an assessor is to divide the text into units. A C-unit was defined in *Figure 7.3* as:

- an independent grammatical predication consisting of a main clause and its associated subordinate clauses; or
- a grammatically incomplete but elliptical statement or response which conveys a recognisable learner aim; or
- an utterance which conveys a recognisable learner aim whether or not in the form of words.

The whole of Extract A, divided into sequentially numbered units, is now given in Figure 7.6. (In the viability exercise, the experts were given texts already divided into units, to save time.)

Figure 7.6: Mark scheme: text in C-units

01	A1: saloon stock cars I think she [pic 14] will not be very interested in it because she's she's more a
02	A2: yeah
01	A1: a one to to look nice
03	A2: I think this one perhaps - they go away to the height Funskating attraction
04	[reading events list] or er or even the windsurfing it can be ...
05	A1: to windsurf
06	A2: yeh, yeh
07	A1: and from Italy she like the sea
08	A2: sure
09	A1: so it's done.
10	Or what do you think about badger and wildlife watch?
11	no, it's not so
12	[both laugh] ...
13	I think it's the best is the windsurfing cos now for the IceTrax it's not the time not the season so
14	A2: so let's choose a
15	A1: Windyfun
16	A2: yeh
17	A1: I'd like to give her lessons [laughs]
	[pause]
18	A1: I think we've chose now

It will be seen that in this case participants' turns coincide mainly with unit boundaries, the exceptions being A2's 03 & 04; A1's 9,10 & 11; and A1's 17 & 18. This extract illustrates not the difficulty of dividing the text into units (the subject of considerable attention in the development of the mark scheme), but the small size of units as elements in a learner's contribution to the exchange (a comment from the viability exercise, which will be reported on later in this chapter). The only unit open to question is 13, where 'cos' introduces another clause which might have been separated into a new unit.

The second task for the assessor is to allocate to each unit one of the list of 15 interactions arrived at as a result of earlier discussion. This final draft list is given

in *Figure 7.7*. The interactions are the same for all scenarios, and an explanation of what each represents is given alongside it.

Figure 7.7: Mark scheme: categories of interaction

- initiation	*starting off a discussion, a new topic*
- acceptance	*understanding and working with other participants' output*
. implication	*making or taking a meaning beyond the literal*
. empathy	*showing understanding of other participants' feelings*
. interpolation	*oiling the cogs of discussion*
- advance	*making a new contribution to the discussion*
. eloquence	*using words to good effect*
. emotion	*expressing own personal feelings*
. enquiry	*asking for information, for other participants' views*
. offer	*helping with information, with opinion*
- response	*taking up the discussion*
. denial	*disagreeing with facts stated or other participants' views*
. substitution	*replacing own word or phrase with another*
. first aid	*asking for or giving missing language*
- conclusion	*finishing the discussion*

Examples of interactions were given in the mark scheme and now appear in *Figure 7.8*: these are all quotations from the transcripts and are intended to act as illustrations for assessors allocating interactions to C-units.

Figure 7.8: Mark scheme: examples of interactions

- initiation	*what happened was...*
- acceptance	*yeh me too I think...*
. implication	*oh congratulations [sarcasm]*
. empathy	*and is she still a friend of yours?*
. interpolation	*yeah, OK,*
- advance	*and so he gets on a boat...*
. eloquence	*she's a fantastic and very good looking woman*
. emotion	*I was very scared*
. enquiry	*what do you think about it?*
. offer	*(A: he's a very well-known scientific -) B: writer*
- response	*(A: ...he meets another David) B: He meets David the second one*
. denial	*no, I think it's quite impossible*
. substitution	*...trying to find a phone pho telephone box*
. first aid	*A: ...past machine? (T: time machine)*
- conclusion	*and that's the story*

The assessor is then to write his chosen interaction in each unit box beside the text, but leave the box blank if the interaction has been unsuccessful, that is, where the participant has not communicated anything.

The third task for the assessor is to score each unit on a five-point scale in terms of structure and lexis. He is to enter one of the ratings ✓✓ ✓ 0 x xx in each box under the headings s and l. Definitions of structure and lexis for this purpose are given in *Figure 7.9*.

Figure 7.9: Mark scheme: definitions of structure & lexis

> **s** - structure: a judgement of what is appropriate for the context, not only the manipulation of words but the links between them, including form (for example, subordinate clauses, word order to add emphasis, etc) and logic (links such as cause and result - 'because...', or reason and consequence - 'and so...');
>
> ***l*** - lexis: again a judgement on appropriateness, not only in vocabulary, but in overall sense and meaning, whether the learner's thought has been suitably expressed, and particularly in phrases which indicate 'chunking' and collocation, including original uses where they are appropriate rather than merely inaccurate. (For example, working with the scenarios, learners have quite often conveyed humour and emotion.)

The assessor's fourth task is to mark on the same five-point scale the four activities set out at the end of the extract by writing the relevant symbols (again, ✓✓ ✓ 0 x xx) in the boxes given - two for each participant. These activities are selected from those listed as 'particular interactions' in *Figure 7.4*: they are related to the content of the scenario and vary according to the kind of task set and the kind of text it produces. The level of marking is to represent the assessor's judgement of how successful the participants have been as a pair/group in fulfilling the demands of the scenario in each activity.

This account has described in some detail the nature of the mark scheme and has been necessarily somewhat extended and episodic, though in practice - as for the viability exercise - it fits on two sides of an A4 sheet, with a third page for a sample of marked text. The following commentary on its operation in practice should bring all the elements together into a coherent pattern.

7.5 The mark scheme: commentary

For the viability exercise, Extract A was used as an example of a marked exchange: this application of the mark scheme, summarising all that has gone before, is now presented in *Figure 7.10*. C-unit boundaries are represented by horizontal lines, and sequentially numbered; an interaction is written in opposite each unit; the achievements of both participants within the unit are then scored with relevant symbols, participants being identified by boxes in which scores are to be entered;

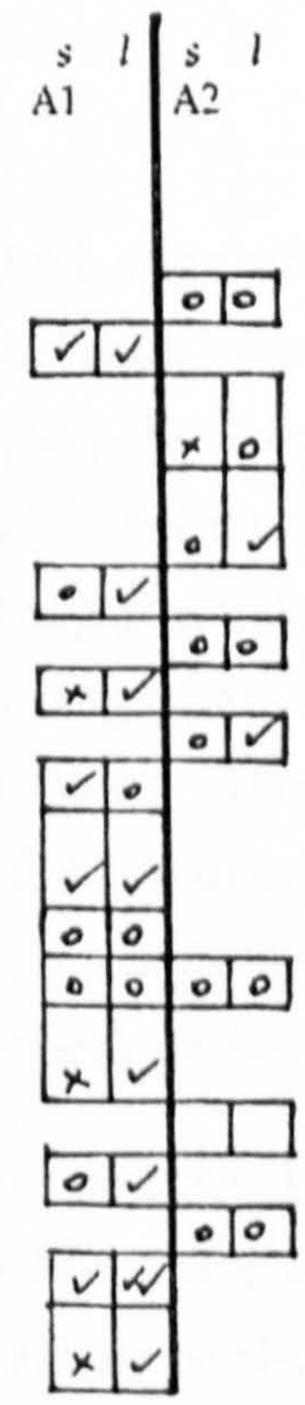
and finally, the activities of each participant are scored, again using boxes. The activities are allocated in pairs to each participant depending on his commitment at this point in the scenario: in this instance, A1 is concerned to discuss the person in the picture he has chosen and to match person with activity; A2 is helping his partner decide but at the same time is interested in moving on to discuss his own choice of picture.

Figure 7.10: Extract A, marked

Extract A

Introduction to Dorset
Context: students, having looked at pictures of people who are going to come to the school as students, are now deciding (task 2) which of several activities (suggested in advertisements) might interest these newcomers.

C-unit	interaction
01 A1: saloon stock cars I think she [pic 14] will not be very interested in it because she's she's more a	
02 A2: yeah	interpolation
01 A1: a one to to look nice	implication
03 A2: I think this one perhaps - they go away to the height Funskating attraction	advance
04 [reading events list] or er or even the windsurfing it can be ...	advance
05 A1: to windsurf	offer
06 A2: yeh, yeh	interpolation
07 A1: and from Italy she like the sea	implication
08 A2: sure	interpolation
09 A1: so it's done.	conclusion
10 Or what do you think about badger and wildlife watch?	initiation
11 no, it's not so	denial
12 [both laugh] ...	implication
13 I think it's the best is the windsurfing cos now for the IceTrax it's not the time not the season so	acceptance
14 A2: so let's choose a	
15 A1: Windyfun	offer
16 A2: yeh	interpolation
17 A1: I'd like to give her lessons [laughs]	implication
[pause]	
18 A1: I think we've chose now	conclusion



activities

- describing	*the characteristics of the person chosen*
- suggesting	*possible matches*
- keeping it short	*to share the discussion with partner*
- encouraging	*keeping the discussion going*

Before an account is given of the viability exercise findings, with general comments on the mark scheme as whole, it might be useful to explore what this example of assessment shows, though it is too short for any general conclusions to be drawn about the qualities of the two participants. According to this assessment, learner

A1 made a greater and more varied contribution to the discussion with 11 of the interactions compared with A2's 9, including one joint comment (laughter - unit 12), as *Table 7.1* shows.

Table 7.1: Extract A, contributions

	A1	*A2*
initiation	1	
offer	2	
implication	3	
acceptance	1	
denial	1	
conclusion	2	
advance		2
interpolation		4
[incomplete]		2
laughter	1	1

The general character of the exchange is that A1 leads and A2 supports, yet A2 not only contributes both examples of *advance* (though A1 may well have contributed others earlier in the interchange, before the extract began), but also keeps the discussion going with *interpolations* such as 'yeah' and 'sure'. A2's suggestions are taken up by A1 with supportive *implications* (line 07 'from Italy she like the sea', line 17 'I'd like to give her lessons') and then adds them together and suggests *conclusions* (lines 09 'so it's done', 18 'I think we've chose now'). The fact that A2 scores zero for interpolations (except 'sure' - line 08) and A1 scores zero for 'no, it's not so' (line 11) do not mean that these contributions count for nothing, merely that they add no information about the participant's structural or lexical skills.

The mechanics of the interchange are represented more directly by the scores entered under *s* and *l* in *Figure 7.10*. A1 scores 4 overall for *s* and 9 for *l*; A2, -1 for *s* and 2 for *l*. The low level of these scores is partly a result of the short length of the extract, but also because A2 is mainly supporting A1's suggestions rather than making any of his own, in spite of the two *advances* discussed above. A2's utterance in line 14 is incomplete, so can score nothing; his support is expressed in 'yeah', which according to the mark scheme does not earn any credit for mechanics (no manipulation, no noticeable lexical content - though it could be argued that it should be awarded ✓ on first use); and laughter is awarded zero, although it indicates a fellow feeling of absurdity in line 12 and appreciation by A1 of his own wit in line 17 (where the ✓✓ for A1 relates to the fact that his 'thought has been suitably expressed' - as suggested in the mark scheme - rather than to his laugh). The numbers of symbols awarded are not totalled because they are intended as a guide to the assessor in writing statements rather than the source of a relatively uninformative score.

The marks given for the four activities listed at the bottom of the extract show that both participants have fulfilled the requirements, but A1 has been particularly adept at making possible matches between the characters represented in the pictures and the activities suggested in the advertisements, adding comments to support his opinions.

7.6 The mark scheme reviewed

In the process of development the mark scheme was applied by the researcher to all eight extracts from the transcripts of the trial series, and for the viability exercise each contributor was allocated a set of extracts for marking. This cumulative experience with it in operation indicated that though it might be feasible in principle, there were problems in putting it into practice effectively. The difficulties were centred on three areas: size of units, definition of categories and scoring of structure and lexis.

Units

It was suspected from the start, and became clear when the first extract was divided into C-units (as demonstrated in *Figure 7.6*), that even with the wider definition for present purposes, the C-unit is too short for practical judgements to be made about the overall quality of a participant's utterances in a scenario. It had been hoped that as experience with the marking built up, some means would be found of combining C-units to make larger chunks, but no such conglomeration opportunities appeared. Further examples of interactions such as the one mentioned earlier in this chapter (*is and we each know each other from several er letters so we know quite a lot of each other* [Extract B] one meaning or two grammatical clauses?) might have led to a more generous interpretation of what a C-unit may embrace. Another source of larger units might be to explore the significance of the number of times a given interaction is allocated successively within a transcript, with a view to identifying conditions for the allocation of more weighty units. For example, if 'offer' is allocated three times in a row, does this imply that this is one large unit, or that there has been some failure in defining categories or applying them?

A compromise solution to this quandary is to relate the division of the text into units more closely to the notion of 'learner aim'. The term 'unit' has been defined (in *Figure 7.3*) as consisting of

- an independent grammatical predication consisting of a main clause and its associated subordinate clauses; or
- a grammatically incomplete but elliptical statement or response which conveys a recognisable learner aim; or
- an utterance which conveys a recognisable learner aim whether or not in the form of words.

The point of interest here is that 'a recognisable learner aim' is mentioned in the

present modifications to the Chaudron/Brock definition (see discussion in §7.2 above), but not in their original definition, which is exclusively grammatical, as reproduced in the first condition above. An alternative approach might be to identify first a learner's aim (his reason for speaking at all), and so to centre the unit on purpose. This might lead to a wider inclusion of related ideas and so a more general allocation of categories. This broader definition is necessarily less precise than the the original, and therefore more open to vagaries in application (ie loss of consistency), but it could result in more realistic divisions of text in terms of assessment, since there is more scope for comment on successes and shortcomings and hence for the writing of helpful statements, for both achievement and diagnostic feedback. This approach is adopted in the exercise reported below using the revised mark scheme.

Definition of categories

There was disagreement between contributors to the viability exercise in the allocation of categories to units, in spite of the definitions and examples given in the mark scheme. In particular, *advance* was found to apply with too little discrimination to many kinds of continuation of the argument. This indicated a need for revision in two directions: both more detailed information about the meaning of the categories and fewer categories to be considered. In the same vein, there were found to be too many categories to be internalised and applied appropriately by assessors within the time available. These difficulties were compounded by the nature of the exercise, which had to cover a range of extracts in order to prove the versatility of the mark scheme, but at the same time could not be expected to take up more than a limited amount of colleagues' time. In circumstances where a mark scheme of this kind was to be used experimentally in a school, more detailed preparation could be envisaged, including agreement trials and further examples on which to apply the scheme, with the added advantage that the classroom material on which the assessments were to be based would be familiar to the assessors.

The solution to both problems seems to be to revise the list by setting it out as a smaller number of 'macro' categories and adding under each of them a number of sub-categories. One weighty macro could be used to label each chunk of text, and a sub-category added only if and when necessary. Thus the apparent ambiguity of *advance* can be avoided by substituting *continuation* and focussing this, as needed, with sub-categories. Further amendments were made to reduce the detail of the listing: the revised version is given in *Figure 7.11*.

Figure 7.11: Categories revised

• initiation	*starting off a discussion, a new topic*
• continuation	*taking the discussion further, elaborating on the topic*
- enquiry	*asking for information, for others' opinions*
- implication	*making or taking a meaning beyond the literal*
- interpolation	*oiling the cogs of discussion*
- offer	*giving new information or opinion, suggesting logical consequence*
• response	*taking up the discussion, agreeing, confirming*
- denial	*disagreeing with facts stated or opinions offered by others*
- first aid	*asking for/giving/finding missing language*
• conclusion	*finishing the discussion*
+	
- emotion/empathy	*expressing own feelings, or understanding others'*

The first category, *initiation*, remains as the starter, but *advance* is abandoned as being a catch-all category, any event following any other being interpretable as an move forward of some kind. Instead, *continuation* is brought in to represent a follow-on after *initiation*, with four possible (but not essential) refinements to the definition: *enquiry*, *implication*, *interpolation* and *offer*, all retaining the same explanations as before, but now set out in alphabetical order for ease of reference. The next heading, *response*, remains as before, with clarifications, if necessary, provided by *denial* and *first aid* (which now includes *substitution*, since this is now to be regarded as an amendment strategy); and *conclusion* also remains as before. The two categories *emotion* and *empathy* are to be added as extra annotations to any unit entry where appropriate, signalled by +e. They were rarely used in the viability exercise because they were considered secondary to a more direct aim, represented by one of the other categories. They both indicate an expression of personal commitment, a giving of self beyond the factual, and there seems in this case no need to differentiate for marking purposes between the speaker's own reactions and his appreciation of others'. The content of *eloquence*, now omitted, is to be regarded as represented by ✓✓ in the marking of lexis, which includes particular skill in the effective use of words.

Structure and lexis

The difficulties with the 'mechanics' of the interchanges and their valuation on the five-point scale are mainly related to levels of operation (eg what can be credited to an 'advanced' student as opposed to an 'intermediate' student?) and the amount of detail to be expected for their assessment (eg does missing 's' in third person present simple count as an error? and if so, how many times must it appear before it is mentioned in a report?).

Both these problems can be relieved to a certain extent by the simple step of reducing the range of marks, which would result in blunter judgements. This development was consonant with the reporting of diagnostic information in the form of statements. The use of a 5-point scale had originally been intended to provide maximum feedback to learners, but the level of detail required was impractical to maintain. Using instead the three-point scale ✓, 0 and x would require only major error or success to be recorded (leaving out of discussion for the moment what this might mean). However, writing statements requires a positive approach, and it was decided to revise the marking so that elements of structure and lexis were positively awarded one of the three options: ✓✓ for 'noteworthy', ✓ for 'acceptable' and x for 'remedial'. This may appear to be no more than a variation on the 0/1/2 unit system of marking, but in practice it starts from the assumption that the learner is building a meaning of some kind - that he has something to say, though may not have the all the means to say it - so that the marking should be concerned with how far he has been successful. Scoring a learner's contribution with 0/1/2 - 'attempt/adequate/apt' - starts from zero, looking at the language used from the listener's point of view: the first of the three judgements implies a listener's response of: 'I didn't understand', the second:'I knew what you meant' and the third 'I appreciate the language you used'. Starting from the other end of the range, with ✓✓, takes the speaker's aim as guide, and entails a positive search for value on the part of the assessor, followed perhaps by some disappointment in reporting a mere ✓ 'acceptable'; 'remedial' then indicates that something can be done to improve the language used. In this way, statements can relate directly to performance. It had always been ruled that incomplete units would not be annotated, and as a further consequence of the blunter approach to marking, it was now decided that unremarkable structure or lexis, ie mechanics within a unit which were not entirely accurate but could not be described as either acceptable (✓) or as justifying the 'remedial' label (x), should be left blank in the columns headed *s* and *l* under achievements. This in effect reduces the spread of marking to a three-point scale with a blank in the middle (✓✓, ✓, , x) instead of a five-point scale centring on zero (✓✓, ✓, 0, x, xx), but since the aim is to provide evidence for direct feedback rather than numerical information, the blank here represents 'no useful information' rather than the 'no score' of 0, and can safely be omitted altogether.

At this point in the development, it was considered that the use of the word *achievements* for the learner's use of structural and lexical elements alone was unhelpful, since the whole of the assessment system of the project was concerned with achievement, that is the learner's progress in language as he experienced it, either in class or outside it, but especially in relation to the contents of a course book. This is in contrast to the assessment of proficiency, that is the learner's current relationship with constructs representing language ability. It was decided to restrict the use of *achievement* to learning, and refer to the learner's use of the

mechanics of language as reflecting his *resources*.

Activities

The revision of the part of the mark scheme concerned with *activities*, that is, participants' pair and group working, was relatively minor. It became clear as the marking was done that the activities for each scenario were not always exclusive to a particular participant: any of the four, though they had been allocated in pairs to each participant, might be evident on occasion in any contribution. A simpler system is to assess across all four for each participant. The principle can be maintained that the same symbols are to be used as for the marking of resources (now ✓✓ ✓ and x) and that these are to be awarded as a result of a scan of interactions. In order to clarify the sources of judgement for the present revision exercise, it was decided to enter on the transcripts the initial letter of the relevant activity against each unit. This helps to explain the workings of the mark scheme, but it is not intended as a requirement for practical applications, where a scan leads directly to the allocation of the symbols, giving an overall view of each participant's contribution to the pair or group work. The interactions to be allocated remain unaltered, as in *List b* of *Figure 7.4* ('particular interactions').

7.7 The revised mark scheme in practice

The mark scheme, now revised in accordance with these findings, was applied to two texts resulting from a further use of the *Books for the library* scenario. The learners were four members of a class preparing for the Certificate in Advanced English (UCLES). The first text is derived from a pair talking about the nature of the books and the second comes from a subsequent group discussion among all four participants about the varying choices of the pairs. These two examples represent different kinds of discussion, the first exploring data with the aim of coming to agreement and the second reporting and justifying to others the agreements arrived at earlier. The data resulting from the marking of these two texts are now to be reported, and a commentary will then follow on these findings.

Results: Text 1

The first transcript is of a text produced by two participants, coded B1 and B2. It provided material for 91 C-units, as defined. The participants were told that they had about three minutes to look at each of a series of books and decide whether they thought it was for adults or for younger students, and what level of language in the book demanded - elementary, intermediate or advanced. It had been suggested that to help them, they should look at the information given for each book on the cover and inside it.

Figure 7.12 gives the first ten units from the transcript.

Figure *7.12*: *Books* text 1, marked

		interaction	*resources*				*activities*	
			s B1	*l* B1	*s* B2	*l* B2	B1	B2
01	B2 I hope it was	-						
02	we've got three more to go	init			✓	✓		k
03	haven't got much time	cont/offer			✓	✓		k
04	[blurb] 'truth to character and situation is the real attraction of the novel should be read by everyone beautiful writing and delicate	init			✓			u
05	I think actually it should be for adults who can appreciate much more than just you know one step at a time	cont/offer			✓	✓		p
	...							
06	B2 what else can we get out of this	init			✓	✓		k
07	B1 wait a minute this is greatest information about this	-						
08	B2 yeah but	resp				✓		
09	B1 [blurb] to catch the spirit of this perfectionist novel	cont		✓			u	
10	B2 but this is for Mermaid Books about the series it is not about this special book this group of books what they represent	resp/denial			✓	✓		p

The interactions, drawn from the revised list (*Figure 7.11*), begin with several *initiations* (02, 04, 06, 07), as both participants contribute to the start of another discussion with reference to a new book. As it progresses, however, there are *continuations* (03, 05, 09) and *responses* (08, 10), and as the argument develops, with *denial* (10).

The participants' resources, recorded in the structure (*s*) and lexis (*l*) columns are unremarkable, with nothing in *s* meriting either 'noteworthy' (✓✓) or 'remedial' (x) comment and nothing beyond routine vocabulary in *l*. There are 6 blanks, representing instances where the wording deserves neither ✓ nor x.

For the purposes of the present discussion the *activities* are to be allocated the initial letter of the relevant category, the four in the case of the Books scenario being u, p, k and r, representing respectively:

- understanding - *what blurb and book imply*
- preference - *saying which book is better & why*
- keeping it short - *to let their partner have their say*
- resolving conflict - *to agree what recommendations to make*

The sample of the text given in *Figure 7.12* illustrates how the revised mark scheme operates; further discussion will now refer to the transcript as a whole, as given in Appendix 7.1.

The contributions of each of the two participants are listed in *Table 7.2*.

Table 7.2: *Books*, text 1: contributions

	B1	B2	totals
[blank]	3	2	4
initiation	5	15	20
continuation/offer	4	15	19
response	11	6	17
continuation	5	11	16
response/denial	1	2	3
continuation/implication	1	2	3
offer	1	1	2
continuation/enquiry	1	1	2
conclusion	0	4	4
first aid	1	2	3
e	0	1	1
Totals	33	62	95

The overall total of contributions adds up to 95 because in 4 cases there were double entries, tallying *first aid* and e at the same time as other interactions. The listing of interactions seems to show at first glance that B2 is the driving force behind the discussion, judging by the number of *initiations* (14 as compared with B1's 6) and *continuation/offers* (B2: 15; B1: 4), and this seems to be confirmed by B1's higher total *of responses* (B1: 11; B2: 7). B1 comes in with *initiations* later on (units 35, 63, 67, 76...). It is B2 however who draws conclusions (28, 51, 61, 90).

The resources deployed by the two participants are uneven, but overall, as can be seen from the transcript (Appendix 7.1), there are few marks of either x or ✓✓ for either structure or lexis: *Figure 7.13* lists these occasions.

Figure 7.13: *Books*, text 1: resources

x

13 B2 I think it's for youngers about all these youngers and confidences ...
24 B2 so it's actually quite intelligent book
31 B2 this is like for everyone who'd like to be remind of childhood
52 B2 and you know for someone who like classics basically like Buddenbrooks or you know
78 B2 OK I would say that this is sort of action story because all this from this serial er this cover what I was saying about it

✓✓

23 B1 and I'd say the problem for us is the real appreciation of the level
48 B2 I would say that this I mean this is psychological action story so you wouldn't give it to children would you? a psychological book
71 B2 and I would say this is intermediate so not very high level ... well one or two hard bits you know

B1 has made no major mistakes but is not flawless, bearing in mind that the revised mark scheme is more generous than the original, since it makes no comment on unremarkable talk, neither good nor bad. B2 (a Polish speaker) has difficulties with articles.

The activities awarded to each participant are listed in *Table 7.3*, on the arbitrary basis of one ✓ per 10.

Table 7.3: *Books*, text 1: activities

	incidence			marks	
	B1	B2	totals	B1	B2
u - understanding	11	21	32	✓	✓✓
p - preference	2	20	22		✓✓
k - keeping it short	0	4	4		
r - resolving conflict	10	4	14	✓	
Totals	23	49	72		

This evidence confirms that, as the table of contributions (*Table 7.2*) demonstrated, B2 is more forthcoming than B1, showing more understanding of what the books are about and especially in giving her opinion on their value for the purposes of the scenario. Some awareness of the time pressure under which they are working is shown by B2 but not by B1; on the other hand B1 is better at bringing the discussion to a close.

Results: Text 2

The second text, with four participants, was expected give a different perspective on the mark scheme, since it represents reporting rather than discussing and agreeing. Each pair of participants was asked to tell the other pair which books they had chosen and whether they were considered suitable for adults or younger learners and at what level. (The full text is again given in Appendix 7.1.)

As a result of experience with applying the revised mark scheme to the first text, units were this time (experimentally) enlarged to include a group of points which could be considered to add up to one meaning. The effect of this was expected to be a blunter, more general assessment of interactions which could be more easily used in the writing of statements. The text provided 45 of these units. No other changes were made in the marking as compared with Text 1.

The contributions of the four learners are listed in *Table 7.4*.

Table 7.4: Books, text 2: contributions

	A1	A2	B1	B2	totals
[blank[	1				1
initiation	3	1		2	6
continuation/offer	1	2		4	7
response	3	5		4	12
continuation	4	3	1	7	15
response/denial		1			1
continuation/implication				1	1
first aid	1	1			2
e	1				1
Totals	14	13	1	18	46

This group discussion is among four participants, two formerly in group A (A1 and A2 in *Table 7.4*) and the others from group B (B1 and B2). It appears that B2 is just as forthcoming with the group as she was in the pair work: she contributes 18 of the 45 units, most of them *continuations*. A1 and A2 offer a greater variety of interactions, but B1 makes very little impact at all.

Figure *7.14: Books*, text 2: resources

x *for structure*

18 B2 ...even if you people would know the language but you have to think what you are reading

24 A2 the vocabulary isn't difficult but the content makes you thinking a lot about

44 B2 it's almost impossible because if if everyone would have different opinion everyone would like...

x *for lexis*

01 A1 ... let me say about one thing we wanted to put the same work as well as you did

25 A2 and really the message and the contents you must read it er many times

34 B2 but it's for entertainment so er instructional book

✓✓ *for structure*
[none]

✓✓ *for lexis*

23 A2 I read this book three times I mean if it was really readable I wouldn't have to read it three times

As can be seen in *Figure 7.14*, the incidence of x and ✓✓ in this text is as sparse as in the first one: 6 'remedial' (x) and one 'noteworthy' (✓✓). There are three 'remedials' under *structure* (18, 24 and 44); and three under *lexis* (01, 25 and 34). The one 'noteworthy' award is for line 23.

The incidence of activities for Text 2 is given in *Table 7.5*.

Table 7.5: *Books*, text 2: activities

	incidence					marks			
	A1	A2	B1	B2	totals	A1	A2	B1	B2
u - understanding	5	8	1	3	17				
p - preference	7	4	1	11	23				✓
k - keeping it short	0	0	0	0	0				
r - resolving conflict	0	0	0	3	3				
Totals	12	12	2	17	43				

This shows a very different pattern from the results for Text 1, for though B2 is still the main contributor to the discussion, she devotes more effort to stating preferences than to understanding. This is a clear reflection of the group's aims, which are now reporting on earlier decisions rather than exploring understanding of the books. None of the group shows any awareness of time constraints: they no longer have the limit of three minutes per book, as recommended for the paired discussion of Text 1. In the same way, since the main aim is reporting, they have little need to resolve conflict. Finally, it should be noted that B1's contribution is very thin, indicating perhaps that she was not as confident as the other three.

Discussion

These variations between participants demonstrate the extent to which the scheme sorts out the individual responses to the same task. In the case of Text 1, both B1 and B2 have the same aims (discussing a book and deciding whether it is suitable for a particular purpose), but respond in different ways, and the mark scheme has succeeded in qualifying their responses in a form which may be represented in an individual assessment report (rather than a score). The revised mark scheme is an improvement on the earlier version in that it is easier to mark and at the same time gives a clearer picture of individual participants' capabilities. The more specific attention to the aims of the speakers in the marking of contributions, and less concern with the form of what they had to say (which had originally controlled the division into C-units), resulted in an increase in the number of words per unit: for Text 2, units averaged about 13 words each, compared with 8 for Text 1. (For the viability exercise, there had been a range of words per unit, averaging at about 5.) This, coupled with the condensation of the list of categories available to the assessor, meant that decisions were easier to make. The change in the marking of resources from a 5-point to a 3-point scale again made the process easier, especially since the blank in the middle eased the marking load even further by reducing the number of entries required. In the same way, allocating all four activities to the contributions of each participant in a pair or group discussion means that the judgements required are simpler and at the same time give a clearer indication of

how participants cooperate in the completion of the task. These blunter judgements, with a consequent loss of precision in the assessment, are probably a more realistic reflection of the process of understanding in informal conversation, where perception of meaning and argument may often be less than entirely accurate.

On the other hand, all these remedies for the surfeit of detail in the original version of the assessment process are likely to reduce the consistency of marking both between one assessor and another and within the work of one assessor. The important question is what level of reliability is required for classroom assessment, and how it is to be achieved, and this was the subject of lengthy discussion in Chapter 3. The answer arrived at there (following the recommendations in Gipps 1994) was to attack the problem from two directions at once, by defining as exactly as possible what criteria were to be used in making the necessary judgements and supplying exemplars of what kinds of decision were to be made. This procedure was followed in the course of the viability exercise, and appeared to be accepted in principle by the experts taking part, their comments being mainly on the criteria set rather than the way in which they were to be implemented. The same principles were followed in drawing up the revised mark scheme.

Some details of the marking have nevertheless thrown up questions of principle. For example, what level of error requires attention (see *Figure 7.13*: verb endings - lines 31 & 52; articles - lines 24 & 78)? In communicative terms, these are minor errors, since the meaning in each case is clear, but they may represent instances of learners' intermediate fossilisation: '...if the underlying system does not evolve, and if communicative effectiveness is achieved, the erroneous exemplar may survive and stabilize, and become a syntactic fossil' (Skehan 1998:61). In this case the question arises as to whether these errors should be reported back to learners so that remedial action may be suggested, even though they do not unduly hamper understanding.

Another uncertainty in the outcome of the marking is the distinction made in the instructions between structure and lexis. For example, with Text 2, line 24 (see *Figure 7.14*), the substitution of 'has' for 'makes' would result in a perfectly acceptable phrase, though one more sophisticated than this participant is likely to attempt ('... the content has you thinking...'). This is a lexical rather than a structural change. Again, with line 25, the need to supply '[to understand]' indicates a lexical error, but adding 'for' at the beginning of the phrase ('for the message and the content...') would be acceptable as a structural amendment. The problem arises from the learner's shortfall in expression, but here again, the ability of a listener/assessor to bridge shortfalls of this kind and make sense of the text as a communication mean that the nature of the feedback to be given is debatable.

Neither of these problems is unique to this project: they are constant occurrences in language classrooms, yet the availability of transcripts and the way in which they have been marked in this instance highlight the difficulty of making decisions about

the formal content of learners' speaking output when the communication of meaning has been constantly reiterated as the most important principle of the present assessment system. All the above argument seems to suggest that users of the system within any particular situation need to agree on criteria in advance so that all can work within the same intentions and so with a high probability of reaching similar decisions. But the overriding consideration is how important consistency is in this relatively informal classroom assessment, where teacher and learners are constantly interacting with each other and a successful outcome is a further advance in the language learning process for all involved.

7.8 Assessment reports

Introduction

The final topic for the project is reporting back to the learner. This is the end product of the series of systems which have been developed as reported in Chapters 4 to 7. It was argued in Chapter 3 that the results of the assessments should be communicated to students in the form of statements, as in Records of Achievement (RoA) schemes in secondary schools in Britain (Broadfoot (ed) (1986) and elsewhere). The benefits of this approach are several:

- it gives information in some detail, encouraging the learner (and incidentally the teacher) by showing what he can now do with his available language, with the option of telling him in addition what areas he needs to revise;
- it provides a record of what the learner has achieved so far, not related to what others have done, but as an individual;
- it helps the learner to see how far he has come rather than how far he has yet to go.

In addition, in the present instance:

- the reports are linked back to learning through the analysis system and so have direct relevance to classroom work;
- the statements result from the learner's engagement with a realistic situation which he has experienced in the environment of the classroom and in collaboration with other learners.

Context of statements

The information given by statements has only limited value for the learner without some indication of how it connects with his overall learning. The mainspring of the present assessment structure is that the learner should be given information about his progress, but 'progress' is a relative term and the reporting needs some mention of both the class work on which the scenario has been based and the context of the learner's contributions on which the statements were generated. To fulfil the first

of these requirements, the report contains as one of its headings a reference to the course work which underlay the construction of the scenario; the second requirement suggests that the statements themselves should include not just a comment on the learner's use of language but also an indication of what he was involved in at the time.

Originating statements

The development of a method for translating the results of scenarios into assessment reports brings up issues similar to those affecting scales which aim to describe levels of achievement. For example, there is a universal need to word statements without ambiguity (or with implications both intended and understood). This ambiguity may be reduced by straight talking to the learner in the simplest possible English, addressing him directly, and making clear the connection between the comments and the learning material on which the scenario was based. Another problem is to achieve a balance between informality and value. A set of statements related to a scenario is not meant to be presented as a formal certificate, but there could be situations in which certification is expected. This expectation is normally to be resisted, for the intention of the project has been to avoid loading assessment with the additional weight of accountability to outside agencies (such as the school, parents, potential employers). If some assessment contract is assumed, it should be an informal one between teachers and learners. The value of the assessment report is then that it becomes an intrinsic part of learning.

Two possible ways of producing statements are either for each teacher/assessor to write them individually for each learner or for a bank of embryo statements to be built up on which assessors can draw, altering and adding to them as necessary. As argued earlier, an essential condition is that statements should derive from actual learner performance rather than from some expert supposition of what speakers say in given circumstances. But the theoretical consequences of this condition are far-reaching if it means that each scenario must have its own set of statements written on the basis of trials and then attached uniquely to that scenario. This is essentially unrealistic for any application of the scheme in a classroom, although it might provide input for a research study on task-based learning. A more practical solution is to collect and categorise statements, all of which are based on actual learner responses to scenarios, but gathered into a bank on which assessors may draw. The list of interactions set up for present purposes (see *Figure 7.11*) will then refer to a variety of contexts and be realised in different ways by different learners; but a fairly generous fit between utterance and statement could be justifiable, provided it gave real information rather than vague generalities. The key to this 'real' information is to relate the content of any statement to the context in which it was generated, by adding to the banked version a link back to the content of the scenario. Thus the comment: *You led the discussion and kept it going*, which could apply to any scenario, would have the context added: *in the group work on deciding which books to recommend*, which would link it specifically to the scenario *Books for*

the Library. This would give immediacy to the comment and at the same time connect with the learner's memory of the scenario as it occurred.

There remains the question of how such a bank of statements is to be set up. The context of the present project is the classroom, with its teacher (or teachers) and learners, and the most likely answer is therefore that the bank would be a collaborative effort among teachers using scenarios in a school, or group of schools.

Guidelines for preparing assessment reports

The principles on which the reports are written, some of which have been discussed earlier in this chapter, may be summarised as follows.

1 Headings are to include name of student, title and date of scenario, reference to course work, comments, signature of teacher/assessor and date of report.

2 Statements should be positively worded, and to improve direct communication, are best written in the second person, addressing the learner as 'you'.

3 A report of weakness should be accompanied wherever possible by suggestions for work on improvement which the learner can undertake for himself.

4 Statements should be derived directly from the work concerned, referring to what the learner has done, rather than what it is assumed (having done it once) he can always do - a prediction which may be contradicted in a lesson following shortly after.

Examples of Assessment Reports

The information from the marking of *Books* Texts 1 and 2, as given earlier in this chapter, is now used as a basis for an illustration of Assessment Reports, one for each of the two learners whose interactions were recorded in the transcripts. The class were working towards the Cambridge Certificate in Advanced English (CAE), and though no detailed analysis as such had been undertaken for lack of time, previous exploration of material published by Cambridge to illustrate its contents (UCLES 1991) showed that a large amount of reading was required within a fairly short time limit (Papers 2 and 3) and that Paper 5, Speaking, demanded paired discussion and reporting, followed by further discussion, as group work. The scenario had previously been used successfully with university-level students who were working with *Headway Advanced* (see Appendix 7.1).

The main findings from the use of the scenario with these learners, as reported in *Figures 7.13* and *7.14* and *Tables 7.2* to *7.5* above, are carried forward to the Assessment Reports, which are given in *Figures 7.15* and *7.16*.

Figure 7.15: Assessment report for B1

The Creative School of English

Student: Barbara Date of assessment 31.04.99

Course: CAE - pair and group discussions - descriptions, comparisons, agreement, reporting back

Scenario: Books for the library

Statements

You were better at answering your partners than at starting the discussion about the next book. You made some good points about the readers you would expect to enjoy the different books, but tended to let the others take the lead. In future, try to take what others say and use it to make your own ideas clearer. Try to be a little more assertive!

The language you used was easily understood by the other students, but you had some difficulties in explaining what you wanted to say, for example about students needing to read books which are not used in class. Listen carefully to English speakers (including on television and radio) and try to hear in detail how they express themselves.

In the pair work you responded helpfully to what your partner said, but in the group you did not contribute very much to the discussion.

Comments

(signed) date

Figure 7.16: Assessment report for B2

The Creative School of English

Student: Betty Date of assessment 31.04.99

Course: CAE - pair and group discussions - descriptions, comparisons, agreement, reporting back

Scenario: Books for the library

Statements

You spoke more than your partners and tended to lead the discussion, especially in the pair work, keeping it going with ideas and suggestions such as what kind of books appeared in the 'Mermaid' series. You were aware that time was important in dealing with each book, and suggested decisions so that you could go on to discuss the next one. You had useful things to say about the readers who would enjoy each book, but you sometimes seemed impatient with your partners and tended to interrupt them: in future you need to listen more attentively to what your partners are saying.

You spoke fluently and generally used language well, but had some difficulty with the use of 'would', especially in conditional sentences. You missed details in grammar, leaving out articles and the ends of verbs: you could try speaking more slowly so that you have time to think carefully about these as you talk. In the pair and group work you contributed well. You were good at introducing new points such as the relationship between ideas and language levels in 'Jonathan Livingston Seagull'. You helped to sum up the discussion so as to bring it to an end.

Comments

(signed) date

These two assessment reports are models rather than detailed examples. They include kinds of comment which may be appropriate in the light of the evidence from the transcripts. The teacher/assessor will have more information about the learners in his class and may add comments which are more individually relevant, and also has pertinent knowledge about the school and the directions of its momentum. The essence of reporting for the present project is that the assessment should be done in a known situation, with the teacher, or group of teachers, in a position to judge what to use or adapt of the systems proposed.

7.9 Conclusion

Summary

This chapter has dealt in some detail with the last step in the sequence which has been reported in previous chapters: after analysis of materials, specification, scenario and practical applications in the classroom comes the assessment of learners' spoken text and the writing of reports. The starting point for this chapter was an investigation of principles on which a suitable marking scheme might be founded, including adaptation of the system for course book analysis, then a return to a new beginning, with a characterisation of learner's aims as they work through a scenario. A mark scheme was devised which assessed three aspects of learner performance: interactions (descriptions of a participant's contributions to the discussion); resources (the language used by a participant to put the interactions into effect); and activities (a participant's involvement in working with others). This scheme was then applied to a set of eight transcripts from the trialling of scenarios (as described in Chapter 6) and a viability study carried out with colleagues as experts to advise on its effectiveness. The scheme was then revised in answer to their comments and applied to transcripts resulting from a further trial with one of the scenarios. Finally, learners' responses were assessed according to the revised mark scheme and assessment reports were written to exemplify how a learner's speaking may be represented in statements so as to give direct feedback on performance.

Evaluation

The marking system as at present organised is capable of providing each learner with information about his operational performance in three important areas which relate directly back to previous classroom work through the systems set up for analysis and specification. This feedback is given in narrative form and includes both positive and negative comments, with suggestions for remedial work where possible.

Taking a still wider perspective, the system has potential for generating useful information about learners' speaking in any classroom where the teaching and learning of language is going on. It is not exclusively attached to the output of scenarios, but may be adapted for the appropriate assessment of language exchanges by learners in whatever context. It could also be useful as a means of investigating language use in classrooms in a different way from classical classroom observation

procedures, which are designed to answer questions about organisational relationships rather than the interests and capabilities of learners, as evidenced by their utterances.

All this is possible because the procedures developed for the assessment of scenario participants are systematic. They go beyond the appraisal of the experienced teacher who may be informally assessing learners continually in class: they are based on an argued theory, detailed development and experience drawn from practical experiment. They are adaptable to a wide range of situations provided only that the principles on which they have been built are maintained, or consciously amended with equally systematic alternatives. In all these uses of the system however it is important that both individual teachers (or other users) and groups should experiment with agreement trials before action is taken on the evidence of results.

The next and final chapter reviews and evaluates what has been reported in the thesis.

Chapter 8

Review

8.0 Introduction

The aim of this chapter is to recapitulate the work done so far, as reported in previous chapters, and to set up some kind of profit and loss account of results.

The original aim of the thesis was to design assessment materials for classroom use which would relate more effectively than traditional forms of testing to current methods of teaching and learning, so that both learner and teacher could arrive at a clearer view of the progress they were making, both severally and together. The link between methodology and assessment meant exploring the implications of a communicative approach for learning as well as for assessment, for example the effects of its concern with context, its emphasis on use and the achievement of practical ends, and hence its interest in the relationship of the individual with language both as learner in the classroom and as user beyond it.

This aim was to be achieved by exploring what it was that teaching materials demanded of learners, with the intention of matching these demands in consequent assessments. The method adopted for the investigation was described as 'empirical', that is to say taking events as found and exploring how they might be characterised so as to suggest ways of dealing with the problems that would arise. Many of these problems were foreseen, but as had also been foreseen, many more were not, emerging as knots in the argument as the work progressed. The empirical approach meant that these could be worked on, and in most instances unravelled, as they appeared.

If classroom assessment was to be as congruent as possible with the preceding learning, the first need was to analyse teaching/learning events in classrooms so that assessments could be designed to follow on more directly from them. This directness was reinforced by adopting a learner's point of view to the demands of the classroom materials by sitting in the learner's place and asking, 'What do I have to *do*?'. The results of the analysis were then to feed into specifications for the design of assessment materials. Over a period of development, these materials evolved into the form of a series of interactions provoked by a live situation in which the learners might reasonably expect to be currently or imminently involved, and in the course of time several of these 'scenarios' were written for trials with learners in language schools. The learners' responses to the scenarios were recorded and transcribed, and a system was set up for judging what the learners had achieved. This mark scheme provided information which could be reported back to

learners in the form of a set of statements.

It is now time to render an account of success and shortfall in the work done, specifically in relation to the following areas of interest:

- the approach adopted
- the systems devised for analysis
- assessment materials
- the trials with learners
- the marking system.

As listed, these represent progressive stages in the thesis and will now be discussed, but in a different order, so that comments may be grouped and considered together under the relevant headings.

8.1 Shortfalls

The most frustrating problem was the division of learner texts, as produced in response to the scenarios, into assessable chunks. There was long debate in the working paper concerned with this topic before the C-unit was decided on as the only feasible chunking system. But in practice it was too short to make practical assessments in the present context. The obvious solution is for participants' texts to be assessed in larger chunks, even at the cost of omitting shorter exchanges and some aspects of mechanics, so as to give more holistic information on the success of interactions. In the longer term, some more empirical approach to chunking might be attempted by consulting a group for expert opinion, asking them to apply defined categories to transcripts without setting unit boundaries, to see what chunking might appear under these conditions. But this too would have its difficulties.

> The teacher who wishes to analyse interaction in his or her own classroom would be well advised not to use ready-made systems of coding and categorizing, but to look for the dynamics, the participation patterns, the work the language is made to do, and to reflect on its relevance to real communicative requirements in the world outside the classroom.
>
> van Lier 1984:168

The 'dynamics, participation patterns and the work the language is made to do' are all represented in the mark scheme as it stands, but van Lier would presumably argue that the procedure is vitiated by the prior categorisation into C-units. Some kind of chunking is needed however for the maintenance of validity and reliability: the difficulty is to find a universal system for the necessary division which depends on the meanings embedded in the text rather than its structural forms.

Another unsolved problem was the inability of the analysis system to cope adequately with structure and lexis as found in course book material. It was obviously necessary to select what was to be entered on the analysis sheets from

the total language given or expected by a task, but it was difficult to decide (in spite of whatever guidelines might be offered by the course book) what was a relatively 'good' or 'advanced' bit of structure or a relatively 'new' or 'original' piece of lexis *for the purposes of the analysis*. The principle of 'as found' remained factual but rather unhelpful in deciding what was to be included or omitted on the analysis sheet, because of the almost unlimited potential for language use in response to the course book's content. Learner responses to a scenario's demands were open and potentially infinite, so that a similar problem appeared in the marking: bearing in mind the level of the learners, what was to be counted as 'good' structure or 'new' lexis *for the purposes of the marking*? But these vast potentialities had been foreseen: the difficulty was that both systems, though potentially useful for research purposes just because of the level of detail pursued, are too fine for easy application in busy classrooms.

The essence of the quandary for both analysis and marking is that attention to detail is essential in a research project if a full investigation of classroom events is to be made. These detailed results are needed to ensure that enough ground has been turned over, but a more general view of the field is needed for practical applications. In the case of the analysis, some progress has been made towards reducing the number of categories, which has helped to make it more cogent and therefore easier to use. In the case of assessment, a blunter system of chunking, accompanied by a revised list of 'macro' contributions has reduced the load for assessors. Further simplification would be helpful, but should be guided by the particular circumstances in which the systems are to be used. The scheme has not yet been used frequently or widely enough to show whether or not it could be a practical addition to (or preferably a replacement for) current testing procedures in schools, and this can only be established by further trials in classrooms.

8.2 Successes

The course book analysis system grew over a series of trials into an applicable method of summarising the demands made by learning material in a range of course books. This may be counted a major success, in spite of the detailed (but reparable) problems with overlaps between categories and the representation of lexis and structure, as discussed above. The computerised version was efficient because it reduced the paperwork to a minimum by offering immediate access to information and examples held in a data base. In the viability exercise, colleagues commented on aspects which were already known to be problematical, but had no difficulty in applying the proposals as a system.

The second part of the analysis, designating clines in the psychological demands made on learners, was successfully implemented by the use of extensive experiment with multiple comparisons of examples drawn from course books. The use of the 'nine-pile' system enabled levels of difficulty to be established in a good example of the empirical approach: what existed was investigated without preconceptions and allocated logically to levels so as to build up into a workable system.

After another extended development period, the assessment materials, as realised in the form of scenarios, were successful in various ways. The intention of the scheme had always been to relate the interactions demanded of learners in the assessment tasks to the circumstances in which they found themselves. This was one of the justifications for the preliminary questionnaire for teachers, which asked for information about the school, its environment and its activities as well as about the course material used. There were several aspects to this relevance of scenarios to learners. For example, the subjects dealt with were connected directly with local events, either within the school (new students, visiting lecturer, school newsletter) or outside it (entertainments, jobs), and as a result learners saw that there could be directly practical applications of the language they had been learning. It was also clear that learners found that the groupings in which they worked (individual, pair, group, class) were familiar reflections of the activities expected by course books and perhaps more importantly, that the language material they found they needed to complete the scenarios did come to mind, even if it was not recognised as having been covered recently in class.

The administrative side of the trials can also be counted a success. The materials were carefully controlled, with a usable layout for scenarios and clear instructions. A practical and effective procedure was also devised for identifying speakers on the recordings, using multiple checks including precoding the note sheet for each learner, noting seating positions, and writing down a loud and significant phrase from each participant to identify voices on the tape recording.

It became apparent in due course, as more scenarios were devised for specific situations, that their social content was potentially universal. Every school has new students, most have a library, many arrange visits out to places of interest, both local and further afield. It seemed likely that many of the scenarios so far written could be readily adapted for use in schools in different parts of Britain, and the principles could be extended to schools in other parts of the world. It may be worth recalling that the first experimental scenario had been discussed, written and tried out, with some success, in Perú. This method of preparing and writing assessment materials might be considered to alleviate the long debated problem of authenticity, since the learners are required not only to use language for a realistic purpose but also to tackle a problem which they can see is directly relevant to their own here-and-now, geographically, socially and personally. This is an end point of what has been discussed earlier as 'individualisation'.

The empirical approach adopted to the research was also successful, as far as it could be carried through. In the abstract, the idea was that avoiding theoretical presuppositions, and yet at the same time adopting a consistently existential stance to the data which appeared, would not only focus on the learner as the centre of the language learning (and assessing) procedure, but would also result in a wide-ranging discussion without setting preset field boundaries which might limit it to

language learning concerns alone. In practice, the result of this approach was a series of 'working papers' which were written to attack problem topics as they appeared, with the intention of drawing in possible solutions from other areas, some parallel, others remote. This process meant that much of the discussion, though of interest (varyingly) in itself, reached conclusions which were rejected as impractical for current purposes. It was a peripatetic (and, it must be admitted, time-consuming) procedure which nevertheless reached mainly practical conclusions.

The end result has justified this approach. Starting and ending in the classroom, the project worked through a series of systems which can be taken up by researchers and teachers if found useful for their own purposes. There is no necessary link between the use of the course book analysis system and the application of that analysis to assessment; no compulsion to undertake the analysis before taking up the suggestions made for the writing of scenarios; no necessity to link the mark scheme with the outcome of scenarios alone. The proposals evolved in the working papers are available as a source of ideas for different applications by researchers, teachers and students in their own fields of interest. For example, scenarios might be devised as learning material without thought of direct assessment; the analysis of course books (or an adaptation of it) could be used to investigate a proposed new course; the mark scheme (or another version of it on the same principles) could be used for the assessment of speaking without reference to the rules for scenarios. The underlying premise is that informed judgement about learners and the language they meet and use is much more informative and helpful in assisting their progress than an abstract score. Since the whole project is based on the assessment of the individual in his current situation, it is only the teacher, and to an extent the learner himself, who can appreciate what are the relevant decisions to be made in terms of curriculum, forms of assessment and the impact of contexts, linguistic, social and geographical.

8.3 Evaluation

It may be useful at this point to consider the achievements of the project in relation to an independent evaluation procedure. Bachman & Palmer (1996:150-155) offer a series of 'questions for the logical evaluation of usefulness' which provides a framework for considering the results of the present project. There are six sections in this list of questions, headed respectively: Reliability, Construct validity, Authenticity, Interactiveness, Impact and Practicality. These areas of concern represent a communicatively-oriented update on the standard three criteria for good tests: reliability, validity and practicality (quoted, with variations, in books on language testing from Lado 1961 to Weir 1993). Bachman & Palmer set out a series of questions under each heading, with space for comments on the 'extent to which' and an 'explanation of how' each of the listed qualities is satisfied. The present project's answers to these questions are given in the following commentary.

Reliability: setting, rubric, input

The setting for scenarios is consistently the classroom, and there is no more variation between the physical conditions in different classrooms than in different examination halls or interview rooms. But there is also a certain consistency in the candidate's attitude to the assessment: for the interview it can be nervousness or fright, which may be argued to represent some realistic interview situations but is not conducive to best performance, whereas for assessments which occur in the normal course of classroom events, such as scenarios, the learners are used to working with others and are ready to enjoy the challenge of new tasks, with the likelihood that their performance will accurately represent their capability.

Scenarios are written according to a format which follows a recognisable, if in detail variable, pattern - individual, pair, group, class - common in course books and therefore familiar, so that rubrics can be short, simple and comprehensible.

The input to scenarios developed into a set of tasks on a standard A4 sheet, with additional material (visuals, texts, realia) as necessary. The sheet is used by participants to make notes at intervals as the scenario proceeds, partly to consolidate the speaking done so far and partly as a reminder for later discussion. In the trials it had a questionnaire on the reverse which asked for participants' opinions: this somewhat elementary form of self-assessment could be developed further as part of the assessment procedure.

Responses to the tasks of the scenarios were intended to vary considerably, because they arose from individual views expressed in answer to deliberately provocative problems. The marking took a positive attitude to this variety by accepting any appropriate response, without attempting to demand any specific structure or lexis. This meant that a suitable mark scheme had to be developed (see below).

Bachman & Palmer suggest that variation in 'the relationship between input and response' is to be avoided. Scenarios are devised and presented so as to take into account variations in local circumstances, including the course book in use and the physical situation of the students. At the same time, they are written in accordance with systematically developed principles which ensure that they represent a consistent approach to the production of assessment materials.

Construct validity: definition, relevance, task, scoring, interpretation.

The intention is to provide an assessment of achievement in classroom activities, and the construct definition for the assessment in this instance is what appeared in the course books used, as sampled for the analysis and defined in the specification derived from it. The purpose of the assessment is to give the students, and the teacher, helpful feedback on how the work in the classroom is progressing: the fact that it is based on classroom work defines the link between construct and purpose. The scoring was done in accordance with a mark scheme which appreciated each learner's engagement with the tasks set, the quality of the language he used and his

contribution to group working, with the intention of providing factual feedback on performance.

In the Bachman & Palmer checklist there are five questions on bias, defined as the way in which setting, rubric, input and so on may cause different takers of the test to perform differently. This they describe as a question of equal opportunity: in the present case the material is written for the learners as a group in a specified context (the course book, and the social and geographical situation) in which all are equally involved.

Authenticity

Bachman & Palmer's concern here is with the correspondence between 'target language use' tasks and test tasks, in other words, the relationship between the test and 'real life'. Scenarios are written for a relevant situation, for example, the arrival of new students at the school, or the choice of books for the school library, and are therefore directly related to possible language uses for the learners. The first few scenarios in the trial series were unquestionably local, referring to entertainments and jobs advertised in a local newspaper; later ones related to actual events in the school as advertised on the notice boards. The content was therefore authentic for participants, involving them in discussions on topics directly arising from their immediate environment.

Interactiveness

Scenarios satisfy Bachman & Palmer's requirements for the 'involvement of test takers' topical knowledge', 'suitability of the test tasks to the personal characteristics of the test takers' and functions 'other than the simple demonstration of language knowledge', since they are devised to meet the demands of a current situation and learners make their own choices in response to tasks, applying their language knowledge to local circumstances. 'Metacognitive strategies' and 'affective schemata', as defined, are involved in following the sequence of events proposed by a scenario and bringing participants' own ideas and choices to the discussions.

Impact

Individuals are involved directly in scenarios to the extent that they are given open situations to which they bring their own interests and skills beyond the use of language alone. Feedback to learners is provided in the comprehensible form of statements on contributions, resources and activities (see Chapter 7). Feedback from learner to teacher on the impressions made by the scenario has been elicited by questionnaire. Interpretations of results and decisions made on them are likely to be talked over in class, but the main outcome is intended to be better information for both learners and teacher on the progress made in their work together. The administrative impact on teachers is likely to be heavy if they are to be individually responsible for analysing classroom activities and writing scenarios,

but the intention is that there should be some sharing of this work among teacher groups within a school. In terms of teaching and learning, the result of the systems should be a clearer understanding of exactly what goes on in classroom learning. The impact of the scheme 'on society and education systems' is more difficult to estimate because it depends on how far the ideas are adopted, experimented with and adapted to the needs of varying learners and institutions.

Practicality

The resources required are mainly, as suggested earlier, teacher interest and teacher time. From the administrative point of view, however, the production of a scenario is simple enough: preparing a single note sheet with back-up material as necessary, to be reproduced in quantity to provide one set for each learner. Feedback on the basis of notes taken while the scenario is progressing has been mentioned but not explored: transcripts were used to demonstrate how the scoring system worked, but if a system of immediate feedback could be devised, it would be more economical and so more practical.

These answers to Bachman and Palmer's questions on the evaluation of usefulness show that the assessment system devised for the present project meets all the demands they propose.

8.4 Outcome

The original aim of the research was to devise a straightforward procedure for writing and using test material which would be effective in the context of communicative teaching and learning. In plain terms, the outcome was to be a comprehensive manual for 'language assessment in the classroom' - a phrase used as the provisional title of the thesis. The end finally achieved has been less simplistic. The exploration of theory which was embarked on to justify the decisions made during the development work revealed areas where universally acknowledged problems, such as the relationship between aspirations and events in the language classroom and the nature of authenticity for the individual learner, remain unresolved in discussions in the literature, but harden as partial solutions into the dubious certainties demanded by daily survival for the teacher.

Instead of a handbook, the thesis now presents a series of working systems:

- for the analysis of course book content (potentially extendable to deal with any classroom events);
- for the design and production of assessment material which relates to the preceding learning and at the same time looks forward to situations which are live and relevant for learners (with potentially universal application);
- for the assessment of interactions produced by learners in response to appropriate materials.

The systems are built on defensible theory and have the potential for further development towards the original aim of providing useful feedback to learners and teachers.

Skehan, in his state of the art paper of 1988, made the following comment about achievement testing.

> ...one of the most serious shortcomings in language testing has been the lack of progress in (or even concern for) the growth and development in proficiency. There has been a corresponding stagnation in the area of achievement testing, and the measurement of learning.
>
> Skehan 1988:219

More than ten years on, the situation now is little better. It is hoped that this thesis may be regarded as a step in the right direction.

References

Abbs B, Ayton A & Freebairn I (1975) *Strategies* (London: Longman)
Alderson J C (1981) 'Report of the discussion on communicative language testing' in Alderson & Hughes (eds)
Alderson J C (1983) 'Who needs jam?' in Hughes & Porter (eds)
Alderson J C (1991) 'Bands and scores' in Alderson & North (eds)
Alderson J C & Hughes A (eds) (1981) *Issues in language testing* ELT documents 111 (London: The British Council)
Alderson J C, Krahnke K S & Stansfield C W (eds) (1987) *Reviews of English language proficiency tests* (Washington DC: TESOL)
Alderson J C & North B (eds) (1991) *Language testing in the 1990s* (London: Macmillan)
Alexander L G (1978) *Mainline beginners A* (London: Longman)
Alexander L G (1988) *Longman English grammar* (London: Longman)
Alexander L G, Stannard Allen W, Close R A & O'Neill R J (1975) *English grammatical structure* (London: Longman)
Allan D & Timmer J (eds) (1996) *Putting the learners in their proper place* (Papers from a symposium of the Testing and Learner Independence Special Interest Groups) (Tankerton, Whitstable: IATEFL)
Allen J P B & Buren P van (eds) (1971) *Chomsky - selected readings* (London: OUP)
Allen J P B & Davies A (eds) (1977) *Testing and experimental methods: The Edinburgh course in Applied Linguistics vol 4* (Oxford: OUP)
Allwright R L (1979) 'Language learning through communication practice' in Brumfit & Johnson (eds)
Allwright R L (1981) 'What do we want teaching materials for?' *ELTJ* 36 1 5-18
Allwright D (1988) *Observation in the language classroom* (London: Longman)
Anastasi A (1968) *Psychological testing* (London: Macmillan)
Apple Computer, Inc (1988) *Apple Macintosh HyperCard* (Cupertino, California: Apple Computer, Inc)
APU (1982/3) see Gorman et al (1984) and Appendix 2.1
ARELS (1967) see ARELS Examinations Trust (nd) and Appendix 2.1
ARELS (1994) *Learn English in Britain with ARELS* (London: Association of Recognised English Language Services Ltd)
ARELS Examinations Trust (nd) *ARELS oral examinations: Rationale, development and methods* (London: AET)
ARELS Examinations Trust (nd) *Key to the Diploma Examination in Spoken English and comprehension AP29* (London: ARELS)
Arnold J (1998) 'Towards more humanistic English teaching' *ELTJ* 52 3 235-241
Austin J L (1976) *How to do things with words* (Oxford: OUP)
Avlp (1985) see Swain M (1985) and Appendix 2.1
Bachman L F (1990) *Fundamental considerations in language testing* (Oxford: OUP)
Bachman L F & Palmer A S (1996) *Language testing in practice: designing and developing useful language tests* (Oxford: OUP)
Baer E R (ed) (1976) *Teaching languages: ideas and guidance for teachers working with adults* (London: BBC)
Baker R (1997) *Classical test theory and item response theory in test analysis.* Special Report No 2, *Language Testing Update* (Lancaster: Centre for Research in Language Education)
Balogh J (1982) *Profile reports for school leavers* (London: Schools Council)
Berne E (1966) *Games people play* (London: Andre Deutsch)
Biggs J B & Collis K F (1982) *Evaluating the quality of learning: the SOLO taxonomy* (London: Academic Press)
Brandes D & Phillips H (1979) *Gamesters' handbook* (London: Hutchinson)
Breen M P (1983) 'Syllabus design: possible future trends: prepared comments' in Johnson & Porter (eds)
Breen M P (1987) 'Learner contributions to task design' in Candlin & Murphy (eds)
Breen M P & Candlin C N (1987) 'Which materials? A consumer's and designer's guide' in Sheldon (ed)
Brims J (1982) *English for negotiating* (Leeds: E J Arnold)
Brindley G (1998) 'Outcomes-based assessment and reporting in language learning programmes: a review of the issues' *Language Testing* 15 1 45-85
The British Council (1978) *Individualisation in language learning: ELT Documents 103* (London: the British Council)
The British Council (1982) *Humanistic approaches: an empirical view: ELT Documents 113* (London: the British Council)
The British Council, University of Cambridge Local Examinations Syndicate & International Development Program of Australian Universities and Colleges (1989) *An introduction to IELTS* (London: BC/UCLES/IDP)
The British Council, University of Cambridge Local Examinations Syndicate & International Development Program of Australian Universities and Colleges (1997) *International English Language Testing System: The IELTS handbook* (Cambridge: UCLES)
Broadfoot P (1980) 'The Scottish pupil profile system' in Burgess & Adams (eds)

Broadfoot P (ed) (1984) *Selection, certification & control* (London: Falmer Press)
Broadfoot P (ed) (1986) *Profiles and records of achievement: a review of issues and practices* (London: Holt, Reinhart & Winston)
Broadfoot P, James M, McMeeking S, Nuttall D & Stierer B (1988) *Records of achievement. Report of the national evaluation of pilot schemes* (London: HMSO)
Broadfoot P, Grant M, James M, Nuttall D & Stierer B (1991) *Records of achievement. Report of the national evaluation of extension work in pilot schemes* (London: Department of Education & Science and the Welsh Office)
Brock C (1986) 'The effects of referential questions on ESL classroom discourse' *TESOL Quarterly* 20 1 47-59
Broughton G, Brumfit C, Flavell R, Hill P & Pincas A (1978) *Teaching English as a foreign language* Second edition (London: Routledge & Kegan Paul)
Brown G & Yule G (1983) *Teaching the spoken language* (Cambridge: CUP)
Brown S (1980) *What do they know? A review of criterion-referenced assessment* (Edinburgh: HMSO)
Bruce G (1969) *Secondary school examinations: facts and commentary* (Oxford: Pergamon Press)
Brumfit C (1979) '"Communicative" language teaching: an educational perspective' in Brumfit & Johnson (eds)
Brumfit C (1981) 'Teaching the "general" student' in Johnson & Morrow (eds)
Brumfit C (1984) 'Theoretical implications of interlanguage studies for language teaching' in Davies et al
Brumfit C (ed) (1986) *The practice of communicative teaching* (Oxford: Pergamon)
Brumfit C (ed) (1995) *The work of the Council of Europe and second language teaching* (London: Macmillan)
Brumfit C J & Johnson K (eds) (1979) *The communicative approach to language teaching* (Oxford: OUP)
Buck G (1994) 'The appropriacy of psychometric measurement models for testing second language listening comprehension' *Language Testing* 11 2 145-170
Bung K (1973) *The foreign language needs of waiters and hotel staff* (mimeo) (Strasbourg: Council of Europe)
Burgess T & Adams E (eds) (1980) *Outcomes of education* (London: Macmillan)
Burniston C (1968) *Creative oral assessment: its scope and stimulus* (Oxford: Pergamon) Second edition 1982 (Southport: English Speaking Board International)
Butterfield T (1989) *Drama through language through drama* (Banbury, Oxon: Kemble Press)
Button D (1978) 'The Rolls-Royce of examinations?' *Education & Training* October 1978
Byrne D & Rixon S (1979) *Communication games* (London: The British Council)
CAE (London) (1986) see University of London School Examinations Board (1986) and Appendix 2.1
Canale M (1983) 'On some dimensions of language proficiency' in Oller (ed)
Canale M (1985) *Proficiency-oriented achievement testing* Paper presented in the Master Lecture Series of the American Council on the Teaching of Foreign Languages
Canale M & Swain M (1980) 'Theoretical bases of communicative approaches to second language teaching and testing' *Applied Linguistics* 1 1 1-47
Candlin C N (1987) 'Towards task-based learning' in Candlin & Murphy (eds)
Candlin C N & Murphy D F (eds) (1987) *Language learning tasks* (London: Prentice Hall)
Carroll B J (1980) *Testing communicative performance* (Oxford: Pergamon)
Carroll B J (1981) 'Specifications for an English Language Testing Service' in Alderson & Hughes (eds)
Carroll B J & Hall P J (1985) *Make your own language tests* (Oxford: Pergamon)
Carroll B J & West R (1989) *ESU Framework* (London: Longman)
Carroll J B (1968) 'The psychology of language testing' in Davies A (ed)
Carter G & Thomas H (1986) '"Dear Brown Eyes": Experiential learning in a project-oriented approach' *ELTJ* 40 3 196-204
Carter R (1987) 'Vocabulary and second/foreign language teaching' *Language teaching* 20 1 3-16
CCSE (1981) see RSA (nd) and Appendix 2.1
Chalhoub-Deville M (1995) 'Deriving oral assessment scales across different tests and rater groups' *Language Testing* 12 1 16-33
Champkin P (1965) *For the employed* (London: Fortune Press)
Chatwin B (1982) *On the Black Hill* (London: Jonathan Cape)
Chaudron C (1988) *Second language classrooms* (Cambridge: CUP)
Clapham C M (1981) 'Reaction to the Carroll paper' in Alderson & Hughes (eds)
Clark J L (1987) *Curriculum renewal in school foreign language learning* (Oxford: OUP)
Clark J L D (ed) (1978) *Direct testing of speaking proficiency: theory and application* (Princeton, N J: Educational Testing Service)
Clarkson P (1989) *Gestalt counselling in action* (London: Sage)
Cohen A (1980) *Testing language ability in the classroom* (Rowley, Mass: Newbury House)
Cook G & Seidlhofer B (eds) (1995) *Principle and practice in applied linguistics* (Oxford: OUP)
Corder S P (1973) *Introducing applied linguistics* (Harmondsworth: Penguin)
Corder S P (1980) 'The teacher's contribution to the language learner's linguistic development' *Modern Languages in Scotland* 20 82-87
Croll P (1986) *Systematic classroom observation* (London: The Falmer Press)
Crookall D (1984) 'The use of non-ELT simulations' *ELTJ* 38 4 (1984) 262-273
CSMS Mathematics Team (1980) *Children's understanding of mathematics: 11 to 16* (London: John Murray)
Cunningsworth A (1984) *Evaluating and selecting EFL teaching materials* (London: Heinemann)
Dakin J (1973) *The language laboratory and language learning* (London: Longman)
Davidson J M C (1973) '"A common system of examination at 16+": some reactions to the School Council Bulletin

No 23' *Modern Languages* 54 1 14-22
Davies A (1977 a) 'Introduction' in Allen & Davies (eds)
Davies A (1977 b) 'The construction of language tests' in Allen & Davies (eds)
Davies A (1978) 'Language testing' Survey article in *Language teaching & linguistics: abstracts* 11 145-59 & 215-231
Davies A (1981) 'Reaction to the Palmer & Bachman and the Vollmer papers' in Alderson & Hughes (eds)
Davies A (1986) 'Indirect ESP testing: old innovations' in Portal (ed)
Davies A (1988) 'Operationalizing uncertainty in language testing: an argument in favour of content validity' *Language Testing* 5 1 32-48 (also in de Jong & Stevenson A (eds)(1990)
Davies A (1997) 'Demands of being professional in language testing' *Language Testing* 14 3 328-339
Davies A (ed) (1968) *Language testing symposium* (London: OUP)
Davies A, Criper C & Howatt A P R (1984) *Interlanguage* (Edinburgh: Edinburgh University Press)
DEIC (1994) see Institute of Linguists (1998) and Appendix 2.1
de Jong H A L & Stevenson D (1990) *Individualizing the assessment of language abilities* (Clevedon: Multilingual Matters)
Department of Education & Science and the Welsh Office (1984) *Records of achievement: a statement of policy (London: DES)*
Department of Education & Science (1988) *Task group on assessment and testing: a report* ('TEGAT report') (London: DES)
Dickinson L (1981) "'Have you got Mr Bun the baker?" Problems and solutions in the use of games, role play, and simulations' *ELTJ* 35 4 381-384
Dickinson L (1987) *Self-instruction in language learning* (Cambridge: CUP)
Di Pietro R J (1987) *Strategic interaction: learning languages through scenarios* (Cambridge: CUP)
Doff A & Jones C (1994) *Language in use* (Cambridge: CUP)
Doff A , Jones C & Mitchell K (1983) *Meanings into words Teacher's book* (Cambridge: CUP)
Douglas D (1994) 'Quantity and quality in speaking test performance' *Language Testing* 11 2 125-144
Ebel R L (1972) *Essentials of educational measurement* (Englewood Cliffs, NJ: Prentice Hall)
Edge J (1984) 'Structuring the information gap' *ELTJ* 38 4 (1984) 256-261
Edge J (1987) Letter to the Editor, *ELTJ* 41 4 308
Ek J A van (1973) 'The "Threshold level" in a unit/credit system' in Trim et al
Ek J A van (1975) *The threshold level* (Strasbourg: Council of Europe; with acknowledgement of contribution of L G Alexander, Oxford: Pergamon 1980)
Ek J A van & Trim J L M (1991) *Threshold level 1990* (Strasbourg: Council of Europe Press)
Ek J A van & Trim J L M (1991) *Waystage 1990* (Strasbourg: Council of Europe Press)
Ellis G (1996) 'How culturally appropriate is the communicative approach?' *ELTJ* 50 213-218
Ellis R (1982) 'Informal and formal approaches to communicative language teaching' *ELTJ* 36 2 73-81
Ellis R (1990) *Instructed second language acquisition* (Oxford: Blackwell)
ELTS (1980) see UCLES/BC (nd) and Appendix 2.1
English Language Teaching Development Unit(1976) *Stages of attainment scale and test battery* (Bicester: ELTDU/OUP)
English Speaking Board (1994/5) *Syllabuses for English as an acquired language* (Southport: ESB)
ESB (1953) see English Speaking Board (1994/5) and Appendix 2.1
Færch C & Kasper G (eds) (1987) *Introspection in second language research* (Clevedon: Multilingual Matters)
FCE (1996) see University of Cambridge Local Examinations Syndicate (1995) and Appendix 2.1
Foster P (1993) 'Discoursal outcomes of small group work in an EFL classroom' *Thames Valley University Working Papers in English Language Teaching* 2 1-32
Foster P & Skehan P (1996) 'The influence of planning on performance in task-based learning' *Studies in Second Language Acquisition* 18 299-324
Fried-Booth D, Willis J & Willis D (1989) *Collins Cobuild English Course tests* (London: Collins)
FSI (1956) see Jones R L (1979) and Appendix 2.1
Fulcher G (1987) 'Tests of oral performance: the need for data-based criteria' *ELTJ* 41 4 287-291
Fulcher G (1990) Review of Underhill (1987) *ELTJ* 44 1 80-1
Fulcher G (1996) 'Does thick description lead to smart tests? a data-based approach to rating scale construction' *Language Testing* 13 2 208-238
Gadd N (1998) 'Towards less humanistic English teaching' *ELTJ* 52 3 223-233, with 'Reply to Jane Arnold' 243-4
Garman M (1990) *Psycholinguistics* (Cambridge: CUP)
Geddes M & Sturtridge G (1979) *Listening links* (London: Heinemann)
Geddes M & Sturtridge G (1982) *Reading links* (London: Heinemann)
Geddes M & Sturtridge G (eds) (1982) *Individualisation* (London: Modern English Publications)
George S E (1981) 'Towards an advanced interaction course' *British Language Teaching Journal* 19 2 82-8
Giglioli P P (ed) (1972) *Language and social context* (Harmondsworth: Penguin)
Gipps C V (1994) *Beyond testing* (London: The Falmer Press)
Goacher B & Reid M I (1983) *School reports to parents* (Windsor: NFER-Nelson)
Golebiowska A (1990) *Getting students to talk* (London: Prentice Hall)
Gonzales R D (1987) 'Ilyin oral interview' in Alderson et al
Gorman T P & Brooks G (1986) 'Assessing oracy' in Portal (ed)
Gorman T P, White J & Brooks G (1984) *Language performance in schools: 1982 secondary survey report* (London: DES)
Gouin F (1893) *The art of teaching and studying languages* transl. H Swan and V Bétis (London: George Philip)
Greene J (1972) *Psycholinguistics* (Harmonsdworth: Penguin)

Grellet F (1981) *Developing reading skills* (Cambridge: CUP)
Grellet F, Maley A & Welsing W (1982) *Quartet* (Oxford: OUP)
Groot P J M & Harrison A (1979) *A specimen test for Threshold level English* (Strasbourg: Council of Europe)
Groot P J M & Harrison A (1979) *A specimen test for Threshold level English: Manual* (Strasbourg: Council of Europe)
Halliday M A K (1973) *Explorations in the functions of language* (London: Edward Arnold)
Halliday M A K, McIntosh A & Strevens P (1964) *The linguistic sciences and language teaching* (London: Longman)
Hardy T (1886) *The Mayor of Casterbridge* (London: Macmillan)
Harlen W, Darwin A & Murphy M (1977) *Match and mismatch* (Edinburgh: Oliver & Boyd)
Harris D P (1969) *Testing English as a second language* (New York: McGraw-Hill)
Harrison A (1975) *Prose translation in French* (Unpublished MPhil thesis, University of London)
Harrison A (1979a) *Institute of Linguists examination review 1978* (London: Institute of Linguists)
Harrison A (1979b) *Techniques for evaluating a learner's ability to apply threshold level proficiency to everyday communication* Document DECS/EES (70) 77 (Strasbourg: Council of Europe)
Harrison A (1980a) 'Current views on the Institute's examinations' *The Incorporated Linguist* 19 2 38-42
Harrison A (1980b) *Diplomatic Service language allowance examinations: (i) approaches to the revision of syllabuses and (ii) syllabus content and operational requirements* (unpublished reports: Foreign & Commonwealth Office)
Harrison A (1982a) *Review of graded tests* Schools Council Bulletin 41 (London: Methuen Educational)
Harrison A (1982b) 'Student-centred testing: assessing communication in progress' *Levende Talen* 372 401-410 (reprinted in Willems & Riley (eds) 1984)
Harrison A (1983a) 'Communicative testing: jam tomorrow?' in Hughes & Porter (eds)
Harrison A (1983b) *A language testing handbook* London: Macmillan)
Harrison A (1985) 'Graded Assessment' in *E206 Block 4 Supplementary Readings* (Milton Keynes: The Open University)
Harrison A (1986) 'Assessing text in action' in Portal (ed)
Harrison A (1991) 'Language assessment as theatre: ten years of communicative assessment' in Alderson & North (eds)
Hawthorne L (1997) 'The political dimension of English language testing in Australia' *Language Testing* 14 3 248-260
Heaton J B (1975) *Writing English language tests* (London: Longman) New edition 1988
Heaton J B (1990) *Classroom testing* (London: Longman)
Herbert D & Sturtridge G (1979) *Simulations* ELT Guide 2 (London: The British Council)
Heyworth F (1984) *Report on workshop 7b: Task-based assessment of communicative performance* (Strasbourg: Council of Europe)
Hicks D, Poté M, Esnot A & Wright D (1979) *A case for English* (Cambridge: CUP)
Hindmarsh R X (1980) *Cambridge English lexicon* (Cambridge: CUP)
HKCEE (1996) see King R (1979) and Appendix 2.1
Hodgson F M (1955) *Learning modern languages* (London: Routledge & Kegan Paul)
Howatt A P R (1984) *A history of English language teaching* (Oxford: OUP)
Hudson L (1967) *Contrary imaginations* (Harmondsworth: Penguin)
Hughes A (1981) 'Reaction to the Palmer & Bachman and the Vollmer papers' in Alderson & Hughes (eds)
Hughes A (1989) *Testing for language teachers* (Cambridge: CUP)
Hughes A & Porter D (eds) (1983) *Current developments in language testing* (London: Academic Press)
Hutchinson M & Young C (1962) *Educating the intelligent* (Harmondsworth: Penguin)
Hutchinson T (1997) *Lifelines intermediate* (Oxford: OUP)
Hutchinson T & Torres E (1994) 'The textbook as an agent of change' *ELTJ* 48 4 315-328
Hymes D (1972) 'On communicative competence' in Pride & Holmes (ed) (1972)
IELTS (1980) see University of Cambridge/The British Council (nd) and Appendix 2.1
Ilyin D (1976) Ilyin Oral Interview Reviewed by R D Gonzalez in Alderson et al (eds)
Ingram E (1977) 'Basic concepts in testing' in Allen & Davies (eds)
Institute of Linguists (1988) *Examinations in languages for international communication: Syllabus effective from 1990* (London: Institute of Linguists)
Institute of Linguists (1993) *Diploma in English for international communication: Syllabus 1994* (London: Institute of Linguists)
Institute of Linguists (1996) *Diploma in English for international communication: Modules I-III & background paper* (London: Institute of Linguists)
Israel Matric (1986) see Shohamy et al (1986) and Appendix 2.1
Jafarpur A (1987) 'The short-cut technique: an alternative for testing reading comprehension' *Language Testing* 4 2 195-220
James C V & Rouve S (1973) *Survey of curricula and performance in Modern Languages* (London: CILT)
Jespersen O (1904) *How to teach a foreign language* (London: Geo Allen & Unwin)
Johns Lewis C (1981) 'Testing communicative competence' *MALS Journal* 6 19-61
Johnson K (1982) *Communicative syllabus design and methodology* (Oxford: Pergamon)
Johnson K & Morrow K (1979) *Approaches* (Cambridge: CUP)
Johnson K & Morrow K (eds) (1981) *Communication in the classroom* (London: Longman)
Johnson K & Porter D (1983) *Perspectives in communicative language teaching* (London: Academic Press)
Jones F R (1991) 'Classroom riot: design features, language output and topic in simulations and other communicative free-stage activities' *System* 19 3 151-69

Jones G (1987) Letter to the Editor, *ELTJ* 41 4 307
Jones K (1980) *Simulations - a handbook for teachers* (London: Kogan Page)
Jones K (1982) *Simulations in language teaching* (Cambridge: CUP)
Jones K (1985) *Designing your own simulations* (London: Methuen)
Jones L (1983) *Eight simulations: controller's book & participant's book* (Cambridge: CUP)
Jones R L (1979) 'The FSI oral interview' in Spolsky (ed)
Jones R L (1985) 'Some basic considerations in testing oral proficiency' in Lee et al (eds)
Karavas-Doukas (1996)'Using attitude scales to investigate teachers' attitudes to the communicative approach' *ELTJ* 50 187-196
Kelly L G (1969) *25 Centuries of language teaching* (Rowley, Mass: Newbury House)
Kenny A (1975) *Wittgenstein* (Harmondsworth: Penguin)
King R (1997) *Can public examinations have a positive backwash effect on classroom language?* (Presentation at IATEFL Conference April 1997)
Kingsbury R, O'Dell F & Wellman G (1992) *CAE Advantage* (London: Longman)
Kirkpatrick B (1987) *Roget's Thesaurus pf English words and phrases* (Harlow: Longman)
Klein-Braley C (1985) 'A cloze-up on the C-Test: a study in the construct validation of authentic tests' *Language Testing* 2 1 76-104
Klein-Braley C & Raatz U (1984) 'A survey of research on the C-Test' *Language Testing* 1 2 134-146
Klippel F (1984) *Keep talking* (Cambridge: CUP)
Krashen S D (1982) *Principles and practice in second language acquisition* (Oxford: Pergamon)
Lado R (1961) *Language testing* (London: Longman)
Lazaraton A (1996) 'Interlocutor support in oral proficiency interviews: the case of CASE' *Language Testing* 13 2 151-172
Lee Y P, Fok A C Y Y, Lord R & Low G (eds) (1985) *New directions in language testing* (Oxford Pergamon)
Lee W R (1965) *Language teaching games and contests* (Oxford: OUP) Second edition 1979
Lewis D G (1972) *Statistical methods in education* (London: University of London Press)
Lier L van (1984) 'Analysing interaction in second language classrooms' *ELTJ* 38 3 (1984) 160-169
London Examinations International (1996) *Certificate of attainment in English: syllabus for 1998 onwards* (London: LEI)
Longman (1997) *English language teaching* (catalogue for 1977)
Lynch B K (1997) 'In search of the ethical test' *Language Testing* 14 3 315-327
Lynch M (1977) *It's your choice* (London: Edward Arnold)
McDonough J & Shaw C (1993) *Materials and methods in ELT* (Oxford: Blackwell)
Mackey W F (1965) *Language teaching analysis* (London: Longman)
McLaughlin B (1987) *Theories of second-language learning* (London: Edward Arnold)
McNamara T F (1995) 'Modeling performance: opening Pandora's box' *Appl Ling* 16 2 159-179
McNamara T (1998) 'Policy and social considerations in language assessment' *Annual Review of Applied Linguistics* 18 304-319
Maley A (1983) 'New lamps for old: realism and surrealism in foreign language teaching' *ELTJ* 37 4 295-303
Maley A (1986) '"A rose is a rose" or is it?: can communicative competence be taught?' in Brumfit (ed)
Maley A & Duff A (1978) *Drama techniques in language learning* (Cambridge: CUP)
Maley A, Duff A & Grellet F (1980) *The mind's eye* (Cambridge: CUP)
Maley A & Grellet F (1981) *Mind matters* (Cambridge: CUP)
Meara P (1980) 'Vocabulary acquisition: a neglected aspect of language learning' *Language teaching and linguistics: abstracts* 13 4 221-246
Medgyes P (1986) 'Queries from a communicative teacher' *ELTJ* 40 2 107-112
Mills D, Doust T, Zodéougan B, Tomalin B & Willis P (1986) *English for French-speaking Africa* (Paris: Armand Colin - Longman)
Ministère de l'Education Nationale (1972) *Le français fondamental (Premier degré)* 4th edition (Paris: Institut Pédagogique National)
Moller A (1981) 'Reaction to Morrow's "Communicative language testing: revolution or evolution?"' in Alderson & Hughes (eds)
Montgomery R J (1965) *Examinations* (London: Longmans)
Morrow K (1977) *Techniques of evaluation for a notional syllabus* (London: RSA)
Morrow K E (1979) 'Communicative language testing: revolution or evolution?' in Brumfit & Johnson (eds) 1979.
Morrow K E (1981) 'Principles of communicative methodology' in Johnson & Morrow (eds)
Morrow K E (1983) 'Some comments on issues' in Hughes & Porter (eds)
Morrow K & Johnson K (1979) *Communicate* (Cambridge: CUP)
Morrow K & Schocker M (1987) 'Using texts in a communictive approach' *ELTJ* 41 4 248-256
Moskowitz (1978) *Caring and sharing in the foreign language class* (Boston, Mass: Heinle & Heinle)
Moys A, Harding A, Page B & Printon V J (1980) *Modern languages at sixteen plus: a critical analysis* (London: CILT)
Munby J (1978) *Communicative syllabus design* (Cambridge: CUP)
Munby S (1989) 'The month records died' *Times Educational Supplement* 29 September 1989
Nicholls S, O'Shea P & Yeadon T (1978) *English alive* 2 (London: Edward Arnold)
North B (1995) *The development of a common framework scale of language proficiency* (unpublished PhD thesis, Thames Valley University)
Nunan D (1988) *Syllabus design* (Oxford: OUP)

Nunan D (1989) *Designing tasks for the communicative classroom* (Cambridge: CUP)
Oller J W (1978) 'Pragmatics and language testing' in Spolsky (ed)
Oller J W (1979) *Language tests at school* (London: Longman)
Oller J W (ed) (1983) *Issues in language testing research* (Rowley, Mass: Newbury House)
Oller J W & Streiff V (1975) 'Dictation: a test of grammar-based expectancies' *ELTJ* 30 1 25-36
OMLAC (1978) see Oxfordshire Modern Languages Advisory Committee (1978) and Appendix 2.1
O'Neill R (1982) 'Why use textbooks?' *ELTJ* 36 2 104-111
O'Neill R (ed) (1987) *The Longman structural readers handbook* (London: Longman)
O'Neill T & Snow P (1977) *Crescent English course* (Beirut: English Language Teaching for the Arab World/OUP SARL)
O'Neill T & Snow P (1989) *Crescent English course 9A* (Beirut: English Language Teaching for the Arab World/OUP SARL)
Oxfordshire Modern Languages Advisory Committee (1978) *New objectives in modern language teaching* (Oxford: OMLAC)
Oxfordshire Modern Languages Advisory Committee (1981) *New objectives in modern language teaching 2* (Sevenoaks: Hodder & Stoughton)
Oxford University Press (1973) *English for business: the Bellcrest series* (Oxford: OUP)
Page B (1973) 'Another look at examinations' *Audio-Visual Language Journal* 11 2 127-30
Page B (1974) 'An alternative to 16+' *Modern Languages* 55 1 1-5
Page B (1982) *Defined syllabuses: crutch or straitjacket?* Paper presented at BAAL annual meeting, September 1982
Page B & Hewett D (1987) *Languages step by step: graded objectives in the UK* (London: CILT)
Parlett M (1974) 'The new evaluation' *Trends in Education* 34 13-18
Parlett M & Hamilton D (1972) *Evaluation as illumination: a new approach to the study of innovatory programs* Centre for Research in the Educational Sciences, Occasional paper 9 (Edinburgh: University of Edinburgh)
Pearce J (1972) *School Examinations* (London: Collier-Macmillan)
Pennycuick D & Murphy R (1988) *The impact of graded tests* (Lewes: The Falmer Press)
Piaget J (1926) *The language and thought of the child* (London: Routledge & Kegan Paul)
Pienemann M, Johnson M & Brindley G (1988) 'Constructing an acquisiton-based procedure for second-language assessment' *Studies in Second Language Acquistion* 10 2 217-245
Pilliner A E G (1968) 'Subjective and objective testing' in Davies (ed)
Pollitt A (1991) 'Giving students a sporting chance: assessment by counting and by judging' in Alderson & North (eds)
Portal M (1986) 'Methods of testing speaking in the APU French surveys' in Portal (ed)
Portal M (ed) (1986) *Innovations in language testing* (Windsor: NFER/Nelson)
Porter Ladousse G (1983) *Speaking personally* (Cambridge: CUP)
Porter Ladousse G (1987) *Role play* (Oxford: OUP)
Prabhu N S (1987) *Second language pedagogy* (Oxford: OUP)
Pride J B & Holmes J (eds) (1972) *Sociolinguistics* (Harmondsworth: Penguin)
Radley P & Millichip C (1993) *Workout upper intermediate Teacher's book* (Walton-upon-Thames: Nelson)
Rea P M (1978) 'Assessing language as communication' *MALS Journal* Summer 1978 45-78 (University of Birmingham Department of English)
Rea P M (1985) 'Language testing and the communicative language teaching curriculum' in Lee et al (eds)
Rea-Dickins P M (1997) 'So, why do we need relationships with stakeholders in language testing? - a view from the UK' *Language Testing* 14 3 304-314
Rea-Dickins P & Germaine K (1992) *Evaluation* (Oxford: OUP)
Rea-Dickins P & Lwaitama A F (1995) *Evaluation for development in English language teaching* (London: Macmillan)
Reber A S (1985) *Dictionary of psychology* (Harmondsworth, Middx: Penguin)
Richterich R & Chancerel J-L (1978) *Identifying the needs of adults learning a foreign language* (Strasbourg: Council of Europe)
Riley P (ed) (1985) *Discourse and learning* (Harlow: Longman)
Rivers W M (1964) *The psychologist and the foreign-language teacher* (Chicago: University of Chicago Press)
Rixon S (1981) *How to use games in language teaching* (London: Macmillan)
Rockwell C (1998) 'Teaching the "RSA Way"' *IATEFL Issues* 14 3 9
Rogers C (1983) *Freedom to learn for the eighties* (Columbus, Ohio: Merrill)
Ross S (1992) 'Accommodative questions in oral proficiency interviews' *Language Testing* 9 2 173-186
Rowntree R (1977) *Assessing students: how shall we know them?* (London: Harper & Row)
Royal Society of Arts (nd) *Examinations in the communicative use of English as a foreign language: Specifications and specimen papers* (London: RSA)
Royal Society of Arts (1985) *The communicative use of English as a foreign language* (London: RSA)
Sano M, Takahashi M & Yoneyama A (1984) 'Communicative language teaching and local needs *ELTJ* 38 3 170-177
Saville N & Hargreaves P (1999) 'Assessing speaking in the revised FCE' *ELTJ* 53 1 42-51
Scarbrough D R (1981) 'Interaction and the individual' *ELTJ* 35 2 83-89
Schwartz C, Davidson G, Seaton A & Tebbit V (1988) *Chambers English dictionary* (Cambridge: Chambers/CUP)
Shayer M & Adey P (1981) *Towards a science of science teaching* (London: Heinemann)
Sheerin S (1989) *Self-access* (Oxford: OUP)
Sheerin S (1991) 'Self-access' *Language teaching* 23 3 143-157
Sheldon L E (ed) (1987) *ELT textbooks and materials: problems in evaluation and development: ELT Documents 126* (London: MEP in association with the British Council)

Sheldon L E (1988) 'Evaluating ELT textbooks and materials' *ELTJ* 42 4 237-46
Shiels J (1995) 'From"Communication in the language classroom"' in Brumfit (ed)
Shohamy E (1994) 'The validity of direct versus semi-direct oral tests' *Language Testing* 11 2 99-123
Shohamy E & Reves T (1985) 'Authentic language tests: where from and where to?' *Language Testing* 2 1 48-59
Shohamy E, Reves T & Bejarano Y (1986) 'Introducing a new comprehensive test of oral proficiency' *ELTJ* 40 3 212-220
Sinclair J McH (1989) Foreword to Willis & Willis
Skehan P (1984) 'Issues in the testing of English for specific purposes' *Language Testing* 1 2 202-220
Skehan P (1988-9) 'Language testing' *Language Teaching* 21 4 211-221 & 22 1 1-13
Skehan P (1989) *Individual differences in second-language learning* (London: Edward Arnold)
Skehan P (1992) 'Second language acquisition strategies and task-based learning' in Skehan & Wallace (eds)
Skehan P (1996) 'A framework for the implementation of task-based instruction' *Applied Linguistics* 17 1 38-62
Skehan P (1998) *A cognitive approach to language learning* (Oxford: OUP)
Skehan P & Foster P (1997) 'The influence of planning and post-task activities on accuracy and complexity in task-based learning' *Language Teaching Research* 1/3
Skehan P & Wallace C (eds) (1992) *Thames Valley University working papers in English language teaching Volume 1* (London: TVU)
Soars J & Soars L (1986) *Headway intermediate Teacher's book* (Oxford: OUP)
Soars J & Soars L (1989) *Headway advanced Teacher's book* (Oxford: OUP)
Soars L & Soars J (1996) *New headway English course: intermediate Teacher's book* (Oxford: OUP)
Southern Regional Examinations Board (1980) *Schools Council Project: History 13-16 CSE examination: The 1980 written examination* (Southampton: SREB)
Southern Regional Examinations Board (1981) *School Council Project: History 13-16 CSE examination: The 1981 written examination* (Southampton: SREB)
Spada N & Frohlich M (1995) *COLT observation scheme: coding conventions and applications* (Sydney: NCELTR, Macquarie University)
Spolsky B (1997) 'The ethics of gate-keeping tests: what have we learned in 100 years?' *Language Testing* 14 3 242-247
Spolsky B (ed) (1978) *Advances in language testing series 2: Approaches to language testing* (Arlington,Virginia: Center for Applied Linguistics)
Spolsky B (ed) (1979) *Advances in language testing series 1: Some major tests* (Arlington: Center for Applied Linguistics)
Stansbury D R (1974) *Record of personal experience* (South Brent, Devon: RPE Publications)
Stansbury D R (1980) 'The record of personal experience' in Burgess & Adams (eds)
Sternberg R J (1985) *Beyond IQ: a triarchic theory of human intelligence* (Cambridge: CUP)
Stevick E W (1976) *Memory, meaning and method* (Rowley, Mass: Newbury House)
Stevick E W (1980) *Teaching languages: a way and ways* (Rowley, Mass: Newbury House)
Stevick E W (1982) *Teaching and learning languages* (Cambridge: CUP)
Swain M (1985) 'Large scale communicative language testing: a case study' in Lee et al (ed)
Swales T (1979) *Record of personal achievement: an independent evaluation of the Swindon scheme.* Schools Council Pamphlet 16 (London: Schools Council)
Swan M & Walter C (1985) *The Cambridge English course Teacher's book* (Cambridge: CUP)
Taylor J & Walford R (1978) *Learning and the simulation game* (Milton Keynes: Open University Press)
TEEP (1984) see Weir C J (1983) and Appendix 2.1
Thorndike E L & Lorge I (1944) *A teacher's word book of 30,000 words* (New York: University of Columbia Press)
T Level (1979) see Groot P J M & Harrison A (1979) and Appendix 2.1
Tough J (1976) *Listening to children talking* (London: Ward Lock Educational)
Tough J (1979) *Talk for teaching and learning* (London: Ward Lock Educational)
Trim J L M (1978) *Some possible lines of development of an overall structure for a European unit/credit scheme for foreign-language learning by adults* (Strasbourg: Council of Europe), republished as *Developing a unit/credit scheme of adult language learning* (Oxford: Pergamon, 1980)
Trim J L M (1996) 'Modern languages in the Council of Europe' *Language Teaching* 29 2 81-83
Trim J L M, Richterich R, van Ek J A & Wilkins D A (1973) *Systems development in adult language learning* (Strasbourg: Council of Europe)
Trinity College (1989) *Syllabus of grade examinations in spoken English for speakers of other languages, 1990 edition* (London: Trinity College)
Trinity College (1992) *Syllabus of grade examinations in spoken English for speakers of other languages, 1993 edition* (London: Trinity College)
Trinity College (1993) see Trinity College (1992) and Appendix 2.1
Trinity College (1994) *Syllabus of grade examinations in spoken English for speakers of other languages, valid 1995-7* (London: Trinity College)
Tucker P J B (1982) 'Defined syllabuses of linguistic content (DSLC)' *Modern Languages* 63 1 39-42
Underhill A (1989) 'Process in humanistic education' *ELTJ* 43 4 (1989) 250-260
Underhill N (1987) *Testing spoken language* (Cambridge: CUP)
University of Cambridge Local Examinations Syndicate (1991) *Certificate in Advanced English* (Cambridge: UCLES)
University of Cambridge Local Examinations Syndicate (1993-5) *FCE/CPE revision project: Information bulletins 1-10* (Cambridge: UCLES)
University of Cambridge Local Examinations Syndicate (1995) *First Certificate in English: Specifications and sample papers for the revised FCE examination* (Cambridge: UCLES)

University of Cambridge Local Examinations Syndicate & The British Council (nd) *English Language Testing Service: User handbook* (Cambridge: UCLES)
University of Cambridge Local Examinations Syndicate & Royal Society of Arts (1990) *Certificates in communicative skills in English: Teachers' guide* (Cambridge: UCLES)
University of London School Examinations Board (1986) *Graded tests in English: Introductory booklet* (London: ULSEB)
Upshur J A & Turner C E (1995) 'Constructing rating scales for second language tests' *ELTJ* 49 3-12
Ur P (1981) *Discussions that work* (Cambridge: CUP)
Valette R M (1967) *Modern language testing* (New York: Harcourt Brace Jovanovich) Second edition 1977
Vollmer H J (1981) 'Why are we interested in general language proficiency?' and 'Issue or non-issue: General language proficiency revisited' in Alderson & Hughes (eds)
Weir C J (1983) *Identifying the language needs of overseas students in tertiary education in the United Kingdom* (Unpublished PhD thesis, University of London)
Weir C J (1988) *Communicative language testing* (Exeter: University of Exeter, also Hemel Hempstead: Prentice Hall, 1990)
Weir C J (1993) *Understanding and developing language tests* (Hemel Hempstead: Prentice Hall)
Wesche M B (1987) 'Second language performance testing: the Ontario Test of ESL as an example' *Language Testing* 4 1 28-47
West M (1953) *A general service list of English words* (London: Longman)
West R (1994) 'Needs analysis in language teaching' *Language Teaching* 27 1 1-19
White R (1983) 'Curriculum development and English language syllabus design' in Johnson & Porter (eds)
Whitehead A N (1932) *The aims of education* (London: Williams & Norgate, republished Ernest Benn, 1962)
Whitney N (1988) Editorial *ELTJ* 42 4 235-236
Widdowson H G (1978) *Teaching language as communication* (Oxford: OUP)
Widdowson H G (1979) *Explorations in applied linguistics* (Oxford: OUP)
Widdowson H G (1984) *Explorations in applied linguistics 2* (Oxford: OUP)
Wilkins D A (1972) *Linguistics in language teaching* (London: Edward Arnold)
Wilkins D A (1973) 'The linguistic and situational content of the common core in a unit/credit system' in Trim et al
Wilkins D A (1974) *Second language learning and teaching* (London: Edward Arnold)
Wilkins D A (1976a) 'Classroom techniques' in Baer (ed)
Wilkins D A (1976b) *Notional syllabuses* (Oxford: OUP)
Wilkins D A (1983) 'Some issues in communicative language teaching and their relevance to the teaching of languages in secondary schools' in Johnson & Porter (eds)
Willems G M & Riley P (eds) (1984) *Communicative language teaching and the training of foreign language teachers* (Nijmegen: Interstudie Institute for Teacher Education)
Willis J & Willis D (1989) *Collins Cobuild English Course 2 Teacher's Book* (London: Collins)
Wiltshire County Council (1975) *Record of personal achievement: materials* (Trowbridge: Wiltshire CC)
Wright A, Betteridge B & Buckby M (1979) *Games for language learning* (Cambridge: CUP)
Wright D W (1980) 'TEFL and ESP in French higher education: the case study and role-play approach' *System* 8 103-111
Wright D (1981) 'Simulation and gaming in TEFL': Resource list. *Simulations/Games for learning* 11 3 140-143

Appendices

Appendix 2.1

Syllabuses containing communicative elements

The following sixteen language tests, or testing systems, have been selected as illustrations of communicative testing elements in practice. The examples are listed in chronological order.

The first two headings give the most useful reference(s) consulted and an outline of the origins and characteristics of each example. The date is that of the first sitting of the test/examination, or for a long-established organisation, the first sitting of an examination reflecting a new communicative-type syllabus.

The next nine headings are derived from the 'recurrent testing themes' set out in Chapter 2. In each case, information is as far as possible quoted directly from the references, but summarised without quotation marks when necessary, and further comments by the researcher may be added [in square brackets] for clarification. Footnotes point to some apparently questionable claims. Finally, the salient characteristics of the test procedure from a communicative point of view are listed under the heading 'attributes'.

ESB

references

English Speaking Board (1994/5) *Syllabuses for EAL (English as an Acquired Language)*; Burniston (1968, 1982) *Creative oral assessment* (Southport: ESB)

origins & characteristics

1953. Established 'to promote & encourage all aspects of oral communication', as an improvement on the then current concern with elocution and correct speech production in examinations for public speaking and drama. Mainly for L1 speakers, but also 'a finely graded series of assessments for students for whom English is not their mother tongue...from the simplest introductory level Foundation through Intermediate to Advanced'. Oral test only.

authenticity 'to encourage enquiry, experiment, discovery & enthusiasm [which] involves the "reporter" in a variety of preceding practical written work & encourages dialogue with skilled adults in & beyond school & college', personal project as topic for presentation & discussion

context situation 'to give confidence in speaking English in a group situation...a supportive atmosphere is promoted'; 'to provide a meeting point of craft, trade, commerce, science, arts & leisure activities through oral, manual & visual presentation to a participating group'

performance task is to express own enthusiasm, explain, respond to questions on topic/artefact; individual written report from examiner to candidate

context language '[at earlier levels] considerable tolerance towards errors in grammar & pronunciation...the main priority is communication'

needs [no mention of analysis]

integrative 'not only effective oral communication but to heighten aural sensibility so that listening becomes an active disciplined experience'; 'all candidates will be expected to be active members of the listening group, asking questions & joining in discussion'

strategies 'candidates sitting in semicircle with the assessor as part of the group'; 'facilities for displaying candidates' pictures, drawings...must be provided'; 'at this level [Advanced] it is important that candidates are able to use strategies to overcome limitations of grammar & vocabulary'

qualitative/quantitative 'useful for examiner to assume that the candidate is "average" until he proves otherwise'; statements illustrate requirements for A-E on fan-shaped diagram; 'although we do not assess numerically it is useful to think in terms of' percentages for each section of the examination, with project weighted double; total aggregated in terms of 5 categories (pass, good pass, very good pass, credit, distinction) to arrive at final grade [guidelines on how to do this]; not summation of marks to a total; individual report

generalisation presentation and discussion skills regarded as transferable to everyday contexts; moderating: 'a wise examiner goes through marks with the teacher at the end of the day & discusses any anomalies'

attributes

- speaking as a key to personal development
- group assessment
- individual reports
- aggregation of judgements

FSI

references

Clark J L D (ed) (1978) *Direct testing of speaking proficiency*, (Princeton: ETS); Jones R L (1979) 'The oral interview of the Foreign Service Institute' in Spolsky (ed); Jones R L (1985) 'Some basic considerations in testing oral proficiency' in Lee et al (ed)

origins & characteristics

1956. The Foreign Service Institute Oral Interview Test (FSI) began as a means of setting standards in L2 for US Government personnel in post in foreign countries. Oral test, 2 examiners, 6 levels of proficiency. One of the most influential systems for oral examining, constantly quoted as a reference point for subsequent developments.

authenticity interview 'as natural as possible', general chat leading on 'from concrete to abstract', including 'interpreter situation' which can 'lend reality'

context situation interview mode predominates, though 'role play' may be included

performance 2:1 interaction; discussion controlled by examiner

context language gradual increase in level of sophistication required over 10-30 minutes' interview/discussion/roleplay

needs originally, language required in foreign postings, but no needs analysis as such

integrative speaking only, no R or W; 5 factors: accent/grammar/vocabulary/ fluency/comprehension: so assessment of listening as well as speaking

strategies not specifically assessed

qualitative/quantitative checklist of 5 performance factors on a 6-point scale (cf Osgood's (1957) semantic differential), eg 'accent: foreign .. .. native', but with experience, 'intuitive judgement as accurate as the checklist'; score converted into a rating on a 6-point scale of proficiency

generalisation '...to observe the examinee function in the language very much as he would under normal working conditions'

attributes

- aim of conducting a natural conversation, but on incline of difficulty
- marking originally based on attribute scale

ARELS Oral

references

ARELS Examinations Trust (nd), *ARELS oral examinations: Rationale, development and methods* (London: AET); *ARELS Examinations Trust (nd), Key to the Diploma Examination in Spoken English and Comprehension AP29* (London: AET); Button (1978) 'The Rolls-Royce of examinations?' *Education & Training* October 1978.

origins & characteristics

1967. AET set up 'for improvement of skills in the use and comprehension of spoken English'; joined up with and eventually taken over by University of Oxford Delegacy of Local Examinations (which offers complementary written papers). Originally designed for use in language laboratory, but since 1988 infra-red transmitter used with individual headphones. Every tape marked twice, with third (final) judgement in cases of wide disagreement. Mark scheme tailored to each set of questions.

[Preliminary, Higher & Diploma levels differ in detail: Diploma only referred to below]

authenticity 'non-literary, authentic[1] materials and situations are used throughout'; 'as a means of communication rather than an academic exercise'

context situation 'in six connected[2] sections: Free Oral Expression, Social Responses, Intelligible Speech, Aural Comprehension, Sustained Speaking & Oral Accuracy'; 'importance attached to the use of English in everyday, realistic situations'; situations are mainly set up by rubric + sentence targeting content of items, with some longer texts for comprehension (from BBC recordings)

performance 'techniques which show a marked improvement on conventional methods of oral examining'; tape-recorded answers in principle, though a few are written; pass candidate 'can be expected to work in an English-speaking environment without serious difficulties'

context language short context (respond appropriately to spoken stimulus); longer context for comprehension & summary

needs 'whether a candidate can survive in an English-speaking environment'; no analysis reported

integrative specific tests on grammar, vocabulary, idiom (sentences); longer texts assessed on pre-defined aspects of comprehension, & on candidate's summary

strategies none specifically assessed

qualitative/quantitative some objective (1/0) marking, some scales (0/1/2/3 & 1-10); impresssion mark (fail/pass/credit/distinction) awarded at end, before marks totalled (to score/200); criteria carefully laid down in mark schemes, but grading of results implies norm reference

generalisability includes wide-ranging samples of language in use; 30 minutes of candidate speech assessed

notes

1 'authentic': manufactured spoken stimulus in unconnected utterances; candidate reads half of dialogue; candidate makes summary of radio broadcast...

2 'connected': no narrative or contextual connection discernible - only connection is the examination

attributes

- utterances for response in social situations
- 0/1/2 marking on appropriateness of response rather than formal accuracy

OMLAC

references:

OMLAC (1978) *New objectives in Modern Language teaching* (Oxford: OMLAC); OMLAC (1981) *New objectives in Modern Language teaching* 2 (Sevenoaks: Hodder & Stoughton)

origins & characteristics

1975. Project of the Oxfordshire Modern Languages Advisory Committee: 'new approach to the definitions of objectives in modern language teaching'; 'sequence of tests at lower levels [than 16+] graded so as to give even the least able pupil a realistic & worthwhile set of objectives'; Levels 1 & 2; tests only - no teaching materials; first operational use 1977. Early example of local groups in GOML (Graded Objectives in Modern Languages) movement of 1970s/80s.

authenticity test relates to visit to France/Germany by pupil with parents (level 1), including interpreting (indirect, through reading & listening items)[1]; storyline given in English to connect m/c items in French/German or English; public signs & notices represented, also French/German objects (drawn as visuals); level 2 relates to pupil staying with family

context situation storyline links items; visuals with speech balloons give context for speaking test, with any appropriate answer credited; role plays in speaking test at Level 2

performance speaking tests 1:1 with teacher (5 minutes), prompts in English (say... ask...); 'defined syllabus' provided

context language individual m/c items embedded in narrative, so short contexts; no extended texts for either reading or listening

needs 'linguistic objectives should be realistic, taking account of pupils' attitudes & needs'; topics include travel, café-restaurant, shopping...; no needs analysis reported; defined syllabus lists words & phrases for speaking, listening & reading respectively

integrative individual items attached to a storyline; 'since most situations require an understanding of both the spoken & the written word, these two skills are tested in combination[2] ; for practical reasons alone speaking is tested separately'

strategies not assessed

qualitative/quantitative objective items throughout, listening & reading marked 1/0, speaking test 0/1/2; 'we are not aiming to discriminate between pupils but rather to celebrate their achievement'; marking is criterion-referenced if each item is regarded as an achievement, but pass marks for the award of a certificate range from 44% to 50% [= 'mastery'?]; intention is that '10% of the candidates will be awarded a "credit", 80% will pass and that 10% will not be awarded a certificate'

generalisation not an issue - internal school tests for feedback & reward; sets of tests on same pattern drawn up for French & German, extending to Spanish, Italian, Russian & Chinese; exported to other LEAs (eg Gloucestershire)

notes

1 listening: dialogues manufactured; prompts read by teacher

2 combination of reading & listening m/c items not related to real-life situation

attributes

- use of storyline to lend realism & interest
- levels from extending social needs: visit with parents at level 1, to family at level 2
- 'tests are designed to enable pupils to demonstrate positively what they can understand and say, not to discover what has not been learnt.'
- 'the language content of the various levels is relevant to all learners whatever their ability' (though originally devised with the lower abilities in mind)

T level

references

Groot P J M & Harrison A (1979) *A specimen test for Threshold level English* (Strasbourg: Council of Europe); Groot P J M & Harrison A (1979) *A specimen test for Threshold level English: Manual* (Strasbourg: Council of Europe)

origins & characteristics

1979. Designed as an example of a communicative[1] test based on The Threshold level (van Ek 1975) & developed 'pragmatically over a period by trying out successive versions ... & improving it progressively by means of feedback from students & teachers...over a period of nearly four years, the present specimen test has gone through five cycles of tryout and revision...in five European countries'.

authenticity based on situations & texts[2] predicated by T level syllabus (L: dialogue, finding your way, filling in a form...; R: tourist information, letter)

context situation role plays for writing & speaking tests

performance texts + m/c or T/F for listening & reading (1/0), speaking test includes reversal (candidate asks questions of examiner) and the unexpected (deliberate mistakes by examiner), speaking marked positively, 0/1/2

context language texts for listening & reading include implication

needs follows T level syllabus

integrative separate tests of 4 skills

strategies speaking includes visuals & tokens, student repair of examiner's factual error

qualitative/quantitative scores totalled & weighted by skill to produce overall score

generalisation 'only one of the many possible tests which could be developed to assess the extent to which students have attained the objectives set out in"the Threshold Level"'[3]

notes

1 'communicative' in principle only - high proportion of traditional testing techniques

2 scripted L text, doctored/made R texts

3 'attained the objectives' - an achievement (rather than proficiency) test, with specific exclusion of any material beyond the T level syllabus

attributes

- reversal & repair of factual error included in speaking test
- detailed guide for examiner in speaking test, derived from candidate answers in tryouts

IELTS

references

UCLES/BC (nd) *ELTS: User handbook* (Cambridge: UCLES); BC, UCLES & IDP (1989); *An introduction to IELTS* (London: BC, UCLES & IDP); BC, UCLES & IDP (1997) *International English Language Testing System:The IELTS handbook* (Cambridge: UCLES)

origins & characteristics

English Language Testing Service 1980, International English Language Testing System 1989, 'further modifications' 1995. Started in 1977 as ELTS, a joint development by British Council & University of Cambridge Local Examinations Syndicate of testing service 'originally designed as a test for prospective postgraduate students' but involvement of International Development Program of Australian Universities & Colleges (IDP) introduced 'a growing demand from other student groups and receiving institutions, especially in Australia, as well as new developments in testing theory, [which] has resulted in this up-to-date, completely revised and flexible testing system [IELTS]'. This now (1997) assesses 'whether a candidate is ready to study or train in the medium of English at an undergraduate or postgraduate level' (Academic Modules) & 'is suitable for candidates who are going to English speaking countries to complete their Secondary education, to undertake work experience or training programmes not at degree level, or for immigration purposes' (General Training Modules).

authenticity reading texts are 'recognisably appropriate' & 'from magazines, journals, books & newspapers'; Writing tasks: to present information from visual, to discuss a problem...; Speaking tasks: speak at length, role play, talk about future plans

context situation ELTS had Source Booklet related to 5 specific areas of study (+ 'General Academic'), intended 'to test English language skills in use' & referred to for Writing & Speaking tests, but abandoned for IELTS. Current texts for both listening & reading provide extended context, linked into background of study (academic) or social demands (general training); use of visuals; writing for specified audience (tutor or examiner); speak about own situation

performance profile of 9-band scores on 4 skills with overall band average

context language completion tests, discussion & argument in writing & speaking

needs no prior analysis evident[1]

integrative separate tests of 4 skills

strategies Listening & Reading tests demand more than factual understanding; Writing expects organisation of arguments, evaluation of ideas; in Speaking, 'assessment takes into account evidence of communicative strategies'

qualitative/quantitative 'individual module scores are added together and averaged for an Overall Band Score. Each Band corresponds to a descriptive statement giving a summary of the English of a candidate classified at that level.'[2]

generalisability up to 30 minutes of pretesting may occur with live administration so as to provide material for banking; continual information collection & test analyses

notes

1 'needs': ELTS had detailed specifications based on discipline areas (speculative); IELTS has academic/general training slants based on skills presumed necessary

2 'Bands': derivation of band from score not evident

attributes

- use of descriptive bands (which have achieved international currency)
- available on demand: 'Test centres can arrange an IELTS administration at any time, according to local need', so 6-monthly renewal of materials
- specific assessment of communicative functions & strategies in Speaking test

CCSE

references

RSA (nd) *Examinations in the communicative use of English as a foreign language: Specifications and specimen papers* (London: RSA); RSA (1985) *The communicative use of English as a foreign language* (London: RSA); UCLES/RSA (1990) *Certificates in communicative skills in English: teachers' guide* (Cambridge: UCLES)

origins & characteristics

1981. Study commissioned by Royal Society of Arts from Morrow (1977) 'to develop new testing procedures to match recent developments in the communicative teaching of foreign languages, which were partly the result of the work done for the Council of Europe'. RSA examination 1981: The Communicative Use of English as a Foreign Language (CUEFL); administered jointly by RSA & UCLES 1989-90; review 1989 'with a view to up-dating the examinations & broadening their appeal'; Certificates in Communicative Skills in English (CCSE) (UCLES/RSA) from 1990.

authenticity 'tasks ... replicate tasks which in the real world, users of a language might actually carry out'; texts for listening & reading are 'normally authentic...though may be slightly modified'; 'where possible the [reading] texts used in the examination are genuine samples of the appropriate text type reproduced in facsimile from the original publication'

context situation 'tasks set will normally ... be "in-Britain"...[which] provides a plausible context for the use of English'

performance candidates can enter for any or all of 4 'free-standing' certificates (Reading, Writing, Listening, Oral Interaction)[1], at any one of 4 levels (3 levels in CUEFL) described in terms of degrees of skill (accuracy, appropriacy, range, complexity) specified at each level for each area. 'Each test is based upon the performance of a number of tasks...specification of these tasks defines what the candidate must do.' Functions specified in Writing & Oral & pair & group work included in Oral

context language pack of source material for Reading, tape for Listening contains 'genuine samples of the appropriate text type', competencies specified as 'degrees of skill'

needs 'target audience is adult users of English as a non-native language', but no needs analysis reported

integrative 4 'areas' (= skills) reported separately

strategies included in specifications of degrees of skill under 'appropriacy', 'range' & 'flexibility'

qualitative/quantitative 'the basis of passing or failing is the candidate's performance relative to the criteria set out here.'

generalisability promoted by full description of target audience & detail of specifications

notes

1 'areas': but R & L tests 'may involve writing'; in Oral interaction 'input for the tasks may sometimes involve fairly extended pieces of writing [ie reading]'. But 'As a matter of principle, we believe that the only way to ensure that equal weight is given to all areas is to offer a set of independent single-skill tests.'

attributes

- separate certificates for skills
- task-based assessment
- authenticity & relevance of tasks & texts
- advance publication of performances necessary for success

APU

references

Gorman R, White J & Brooks G (1984) *Language performance in secondary schools: 1982 secondary survey report* (London: DES); Gorman T & Brooks G (1986) 'Assessing oracy' in Portal (ed); Portal (1986) 'Methods of testing speaking in the APU French surveys' in Portal (ed)

origin & characteristics

1982/3. Language Monitoring Project set up by the Assessment of Performance Unit of Department of Education & Science. 2 examples:

a) 1982/83. Oracy section of surveys in L1 of 11- & 15-year-olds: 'It was an intrinsic objective of the oracy programme to broaden the range of types of talk tested...beyond... CSE'; 'appropriateness...an important criterion of assessment'; 'main function of talk is to communicate';

b) 1983-85. L2 (French) tests for 13-year-olds: 'communicative in character', 'whole ability range', 'concerns about face validity', 'conventionally reliable test instrument', 'training of speaking assessors...figure[s] prominently in the preparation' for the S tests

[References below to a) only unless b) mentioned]

authenticity realistic tasks (eg discussion over differing details of 2 maps) but visuals simplified 'to reduce the complexity of the maps...& reduce the dependence of the ability to perform the task on previous experience'

context situation extended exchanges on visuals (description, discussion)

performance 'collaborative problem-solving talk'; 9 functions & tasks 'developed interactively'; tasks for pairs (pupil & chosen friend)

context language appropriateness (rather than correctness) assessed in dialogues provoked by situations; b) 3 types of task, all cued in L1, but interlocutor (assessor) responds in L2

needs 'different purposes require different varieties of talk', which are sampled in tasks

integrative 'listening should not be artificially separated from speaking...assessed implicitly by assessing some form of outcome'

strategies 'orientation to listener' includes 'both verbal features such as tempo, pausing and "stabilisers" (ums and ers), & also, where appropriate, such non-verbal features as eye-contact, posture & gesture'; map task shows 'the ability to make inferences from conflicting evidence...willingness to refrain from interrupting...when a difference...might not be critical, coupled with an ability to distinguish such cases from those when clarification was essential'

qualitative/quantitative general impression of 'communicative effectiveness' on scale 1-7; impression of 'orientation to listener' on scale 1-5; subsequent analytic marking on categories (sequential structure, lexical selection, syntax, performance features), consolidation not discussed

generalisation wide range of functions & tasks, surveys with N=10,000

attributes

- assessment of listening through oral interaction ('orientation to listener' strategies)
- interaction between function & task
- appropriateness rather than correctness as criterion of success

TEEP

references

Weir C J (1983) *Identifying the language needs of overseas students in tertiary education in the United Kingdom* (Unpublished PhD thesis: University of London); Weir C J (1988) *Communicative language testing* (Exeter: University of Exeter)

origins & characteristics

1984, now discontinued. Associated Examining Board developed 1978-84 'examination which would assess performance of students in English for Academic Purposes...later called Test in English for Educational Purposes (TEEP).'

authenticity 'performing language tasks relevant to the academic context in which he has to operate'

context situation likely texts & tasks, but not settings; speaking test is with tape (cf ARELS) & requires 'respond to remarks...imagine in a discussion group...tasks on non-verbal (graphic) & verbal/numerical tables'

performance reported in terms of 5 grades

context language source booklets for Papers 1 & 2 provide extended context for candidates 'to demonstrate proficiency in...enabling skills', eg conceptual meaning, distinguishing main idea from supporting detail, skimming, attitudes, inferences...

needs based on research 'ascertaining the language demands made on students in the disciplines most commonly studied by overseas students'

integrative separate assessments of 4 skills (listening, reading & writing in Papers 1 & 2 & speaking in an optional Paper 3), but 'include a more integrated task in each of the Papers, in which reading &/or listening activities lead into a writing task'; results in form of profile of 5 grades (same wording for each skill)[1]

strategies 'to test a candidate's proficiency in the range of enabling skills required to operate successfully in the various study modes'

qualitative/quantitative 'there will be no overall result, since it is felt that different subjects & departments vary'; so profiles; objectives apparently not used as criteria in marking

generalisation 'Two versions of the second component...intended for...Arts, Social, Business, Administrative Studies...[& the other for] Engineering & Science'; extensive trials; detailed specifications

notes

1 derivation of grade from score not described

attributes

- researched needs
- wide range of communicative objectives
- source booklets
- descriptive grades

AvIp

reference

Swain (1985) 'Large-scale communicative language testing: a case study' in Lee et al (1985)

origins & characteristics

1985. Outcome of a project to develop a testing unit, *A vous la parole* (AvIp), for students at the secondary school level in Ontario, as one of a series of testing units 'to be used in province-wide assessments of the communicative performance of immersion students'.

authenticity '12-page student booklet... about two summer employment possibilities for youth';'six communicative tasks commonly required of a native speaker'

context situation booklet as source, tasks derived from areas for student decision

performance assessed in four areas: grammatical, sociolinguistic, discourse, strategic (following Canale & Swain 1980)

context language 'discourse competence involves mastery of how to combine grammatical forms and meanings to achieve a unified spoken or written text in different genres...'

needs 'in informal sessions...project staff explored topics of greatest personal relevance & interest of these students'

integrative 'the central theme is summer employment for youth...provides the central focus for all the tasks which the students are asked to do'; 'the specific tasks...must in their entirety provide the opportunity to use each component of communicative language behaviour'

strategies 'strategic performance' included as a component but 'not scored in written data'

qualitative/quantitative 'scored by mixture of objective counts & subjective judgements', working 'from data gathered during pilot testing to determine what specifically would be scored in each task & what scoring criteria would be used'; 'selection [by factor analysis] of variables from a much larger set, the larger set being determined by the nature of the responses to each task'

generalisation intended for large-scale use, but replication of process (eg by new materials, different levels) not discussed

attributes

- application of communicative constructs in development of practical tests
- use of source booklet for consistency of topic, integration of language use & authenticity of tasks
- mark schemes derived from answers

CAE (London)

references

ULSEB (1986) *Graded Tests in English: Introductory booklet* (London: ULSEB); London Examinations International (1996) *Certificate of Attainment in English: syllabus for 1998 onwards* (London: LEI)

origins & characteristics

1986. Originated at the University of London School Examinations Board as Graded Tests in English (GTE), 'a series of tests at 5 levels of increasing difficulty... achievements for each level described in detail on the certificates...students take the tests in a language laboratory...take part in a sequence of events.' As Certificate of Attainment in English (CAE), 'launched worldwide in 1988'. Language laboratory test discontinued 1993, & oral test (optional) introduced. Syllabus continually revised (eg addition of a Level 6, introduction of writing at Levels 1, 2 & 3), listing of syntax, morphology & lexis...) but retaining principle of 'tasks based on a situation to which the candidate can relate'.

authenticity GTE: fictional but realistic story, scripted dialogues, concocted texts; CAE: 'realistic & practical situations'; from 'realistic rather than authentic speech' in listening at Level 1 to 'realistic or authentic speech' at Level 6

context situation scenario/situation continues throughout tests (including oral in CAE)

performance GTE: 'opportunity to show how much...English they can cope with, ...realistic tasks...personal commitment', positive marking of student utterances & student writing as related to level descriptions of practical (functional) purposes; communication takes priority over accuracy; CAE: 'tasks to which they can apply their language skills', 'allow candidates freedom to show what they can do', 'to enhance motivation by use of a positive marking scheme'[1] , oral is paired interview with interviewer & assessor & includes 'scenario-related questions'

context language GTE: 'students take part in a sequence of events...extended context'; CAE: 'tasks based on a situation to which candidate can relate'

needs GTE: only as suggested by course books[2]; CAE: Level 1: 'for students of all ages who are starting to learn English'; Level 6: 'for those using English for postgraduate study, or regularly in work at the highest professional level'[3]

integrative GTE: test related to cassette recording with booklet containing visuals & (at higher levels) text, skills as required to complete tasks; CAE: 'specific objectives' in terms of 4 skills (LSRW), parts of tests headed 'Listening', 'Reading', 'Writing', 'Exercises involving the structure of discourse' etc; syllabus includes lists of 'syntax, morphology & lexis' per level.

strategies GTE: assessed as part of response to task

qualitative/quantitative results in form of certificate with statement of level on the back, criterion-referenced marking system, but no policy evident on mastery level

generalisability GTE: sampling based on specific course books; CAE: list of recommended published materials available

notes

1 'motivation': students not necessarily aware of how marking system works, so motivation doubtful

2 'course books': GTE syllabus based on content of *Strategies* series (Abbs et al)

3 no evidence of needs analysis for either GTE or CAE

attributes

- use of scenario
- positive marking
- student commitment to completion of story (GTE)

Israel Matriculation

reference

Shohamy E, Reves T & Bejarano Y (1986) 'Introducing a new comprehensive test of oral proficiency' *ELTJ* 40 3 212-220

origins & characteristics

1986. Research project leading to implementation of new oral procedures for the EFL Oral Matriculation test administered by the Ministry of Education in Israel. 'Four tests...a variety of speech styles...needed in various oral communicative situations.' Aim was to replace existing test, & a new test has been introduced (1986) but 'a number of changes have been made' to the tests used in the project. [All references below are to the project tests]

authenticity 'represent a broader range of oral speech styles [than existing matriculation] which would reflect the type of communicative language that students are likely to need in authentic situations'; topics related to possible experience of candidates (eg role-play: 'crossed road on red light...try to persuade him [policeman] to let you off'; group discussion: 'schools in Israel')

context situation role play, discussion with 3 other candidates

performance assessed on 7-point scale on several variables: intelligibility, language produced, interaction [at lower end], expression, interaction & error [at upper end], taking part in interview, role-play, reporting, argument with peers

context language topics familiar to students, discussion in groups of 4, reporting content of Hebrew text, 'expanding the number of speaking interactions...will create positive washback effects'

needs 'adding other speech styles...involves the use of...features...which are part of the language proficiency that students are expected to acquire by the end of Secondary School'

integrative role play, group discussion

strategies interaction a criterion in rating scale; L1 as input

qualitative/quantitative holistic rating scale, 4 - 10, same for all 4 tests[1]

generalisability 'Although these situations do not embrace the whole scope of oral performance, they were considered by a group of local experts in the field to provide a representative sample of useful oral interactions'

notes

1 'scale': but scores summed & used for traditional data analysis

attributes

- range of tasks
- group discussion
- interaction rated
- 'combining educational research with educational policy'

Trinity College London

references

Trinity College (1989) *Syllabus of grade examinations in spoken English for speakers of other languages, 1990 edition*; Trinity College (1992) *Syllabus of grade examinations in spoken English for speakers of other languages, 1993 edition*; Trinity College (1994) *Syllabus of grade examinations in spoken English for speakers of other languages, valid 1995-97* (all published London: Trinity College)

origins & characteristics

1993. New syllabus. 'Founded in 1872...first to devise the system of offering to external candidates progressive examinations based on carefully graded syllabuses... first in musical subjects, then in speech subjects, & in 1937...spoken English for foreign students...introduced'; new syllabus 1968, revisions; latest changes for 1993 'aim to reflect more recent developments in the theory & practice of language teaching, learning & testing'

authenticity 'the learner is engaged in a real conversation with the examiner'; 'the topic (prepared & introduced by the candidate) provides the opportunity to talk about something of personal relevance & interest';

context situation discussion of topic & (at upper levels) text, both prepared by the candidate; individual face to face with examiner for 'a natural flow of conversation'

performance 'Each Grade has specific task & language requirements'; '"candidate can" statements at each Grade level ... are intended as samples which are characteristic of the relevant level of language ability as a whole'

context language 'criteria of assessment' include 'readiness (understanding, fluency, participation), pronunciation (...variety of speech patterns), usage (...to use language appropriately), focus (organising what is said...purposes...strategies)'[1]

needs 'provision is made for all types of learners - from those still at school to those seeking to develop their language skills for professional, academic or other purposes'

integrative 'the topic provides the opportunity to integrate English language skills'

strategies 'communication strategies' as one of criteria in all examinations; included in "can..." statements at higher levels, eg 'within conventions of normal turn-taking' (Intermediate Grade 7), 'rephrase where necessary" (Grade 8), 'use native speaker communication strategies' (Grade 12)

qualitative/quantitative 'assessed throughout on a 5-point scale, A to E. The letters are converted to marks which total 100. Pass level is 65%...Merit 75%...Distinction 85%'

generalisability stages intended to correspond to levels on the ESU scale; finely detailed requirements for each Grade (listed as 'can' statements[2]) should promote consistency between sittings

notes

1 'purposes': but examples include exchanges without information gap, eg 'How do I get to the station? what's the girl in the picture doing?' (both Grade 3)

2 'can statements': but no supporting evidence for placing a statement at a particular level, eg 'state communicative purpose in simple terms' (Grade 5), 'clearly demonstrate the communicative purpose' (Grade 9); 'maintain the flow of communication by using appropriate linking language' (Grade 6), 'maintain conversation in a natural way' (Grade 11).

attributes

- 12 Grades in 4 Stages, all fully detailed in specifications under 4 or 5 heads (depending on Stage)
- 'on demand' throughout the year
- student chooses topic (Grade 4 up) & text (Grade 7 up) & presents them (cf ESB)
- 'can' statements

DEIC

references

Institute of Linguists (1988) *Examinations in languages for international communication: Syllabus effective from 1990*; Institute of Linguists (1993) *Diploma in English for international communication: Syllabus 1994*; Institute of Linguists (1996) *Diploma in English for international communication: Modules I - III & background knowledge paper*

origins & characteristics

1994. Institute's examinations (originating 1911) at 5 levels, starting from low level base (Preliminary certificate) to professional qualification for translators & interpreters (Diploma). Latest syllabus of modern language examinations introduced in 1990 (ELIC), of revised Diploma in English 1994 (DEIC). Currently only Diploma ('= first degree') level available in English on new syllabus (other levels in preparation). Modular, authentic materials, realistic tasks.

authenticity 'materials are drawn from authentic original sources... tasks... correspond as closely as possible to realistic language use...each task has a specific, realistic purpose'

context situation candidate's brief for each module sets task to be carried out, eg (module 1) with materials & discussion with oral examiner (role play), then writing of report for stated audience

performance module I: 'present, provide & elicit information'; module II: 'argue/ debate a case/negotiate an agreement on behalf of an organization'; module III: 'summarize...communicate in writing'; background: write 'knowledge of topics of interest internationally to the English-speaking world'; candidate interacts with texts & examiner to achieve a stated end; grades Pass, Distinction or Fail in each module, on successful completion of each section of the task set according to an assessment checklist - 'candidates are expected to cover the following points' - special requirements laid down for Distinction for each module

context language eg (module I) extensive reading material to inform candidate of topic in form of 'dossier' to read & make notes on 1 hour before test; candidate then 'gives presentation of main points, requests additional information from oral examiner... writes up findings as short report'

needs 'variety of skill areas [which] will...benefit the needs of employment & those wishing to pursue further study & training at university level'; no research reported

integrative 'examination tasks...require candidates to carry out a chain of integrated activities leading to a final product. Adequate performance in all these activities is a prerequisite for task fulfilment'

strategies 'essential for candidates to be able to deal with the many implications of cultural differences'; language strategies essential for successful completions of tasks

qualitative/quantitative criterion-referenced assessment, with stated requirements for pass & distinction

generalisability sampling by topic, with extensive dossier on topical issues; requirements set out in terms of 'assessment focus' (abilities, functions), 'skills involved' & 'steps' (demands of task)

attributes

- original materials
- realistic tasks
- integrated skills
- criterion-referenced

FCE

references

UCLES (1995) *First certificate in English: Specifications and sample papers for the revised FCE examination*; UCLES (1993-5) *FCE/CPE revision project: Information bulletins 1-10*

origins & characteristics

1996. Originally introduced by University of Cambridge Local Examinations Syndicate (as Lower Certificate in English) in 1939, extensively revised & renamed First Certificate in English in 1974, major revision 1984, revision project 1991-1995. 'Most widely-taken in the range of Cambridge examinations in English as a foreign language...with wide international currency'

authenticity 'important aspect of revision...to consider authenticity', 'as relevant as possible to candidates'; Reading paper includes texts which have a similar topic & genre to those of a text which an FCE candidate might expect to meet in non-test contexts...tasks...include ones which engage them in processing and manipulating English in a similar way to that in which they might outside the test situation... select the most appropriate item for a given context, a task they would realistically set themselves...'[1] .

context situation 6 topics (eg travel & tourism, friendship & social contacts) in 3 settings ('in an English speaking country, in non-English speaking countries in which candidates might use English as a lingua franca, in the candidate's own country'); contexts clear for L & R tests, audience stated for W, peer discussion in S

performance pairs in S, 'test focus' specified for each of 5 papers, eg 'assessment of candidates' ability to understand gist, main points, detail...write specified text types for a specified audience & purpose...exchange information, express & find out about attitudes & opinions'

context language extended texts set for R & L & expected in W & S

needs 'information on candidates' needs & interests is continually collected & evaluated through Local Secretaries' meetings, the market survey questionnaires, Candidate Information Sheets & feedback given by candidates & centres on the contents of trial tests'

integrative 4 skills tested separately in 4 papers

strategies design of the revised examination...aims to provide...coverage of the language abilities underlying these needs & interests (in reading, writing, language systems, listening & speaking)

qualitative/quantitative apparently scored traditionally as input to norm-referenced system (standard at UCLES)

notes

1 'authenticity': but R texts are not as source either in wording or layout; L texts seem to be scripted speech; 'use of English' based on tests of manipulation of grammar & vocabulary...

attributes

- intentions (audience, purpose) specified
- user-led

HKCEE

reference

King R (1997) *Can public examinations have a positive backwash effect on classroom language?* Presentation at IATEFL Conference April 1997

origins & characteristics

1996. Hong Kong Education Authority (instituted 1979), responsible for Hong Kong Certificate of Education Examination (HKCEE) taken at age 17 (N=120,000). Filter for six-form education. 4 papers: writing, reading, integrated listen/read/write, oral.

authenticity published stimulus material adapted to fit exam paper, realistic tasks in integrated paper, 'highest priority is given to ensuring good face validity'; in oral, 'role should reflect real-life situations'

context situation in writing, 'purpose...play a role & know intended audience'; sequence of events in integrated paper

performance in oral, role play with 2 examiners, then discussion in group of 4

context language 'we test grammatical competence indirectly: ie in context', 'extensive use of passages', redundancy of data in integrative test

needs 'we aim to narrow the gap between the language tasks set...& the kind of tasks that will be faced by a Chinese speaker of English in the HK work/study environment'

integrative 'priority to setting tasks which need the integration of a range of language skills'; in R/L/W test, data file, own Walkman radio, 15-minute broadcast, 4 tasks to be completed in role

strategies 'able to function in a group...appropriate interaction skills...turn-taking strategies... helping to reach a consensus'

qualitative/quantitative [no information]

generalisability problem in converting teachers & students from traditional memorising procedures (learning 'lists of set phrases to cope with situations in oral test')

attributes

- public examination using broadcast listening texts
- integrative paper
- 2 examiners & group of 4 in oral
- logistics: 500-600 examiners over 20 afternoons, examining 600 candidates each [but these figures do not tally with N entries quoted above]

Appendices

Appendix 4.1

Analysis guide:

The guide explains how to apply the system for analysing course book materials which has been devised for this thesis. It uses the computer program HyperCard (© Apple Computer Inc 1988).

It consists of 118 'cards' which are here printed eight to a page. There are three kinds of card, containing:

- administrative instructions about how to work the system;
- guidance on what to enter on the analysis sheets;
- a glossary of the terms used in the analyis.

The procedure is self-explanatory, in that once the program has been initiated, the user needs only to follow the information given on successive cards in order to complete the Analysis. In addition to the program and a computer to run it, he needs a sample of a completed Analysis sheet, blank Analysis sheets for completion, the course book (both student's and teacher's, if existing), and a pencil.

Every card includes 'buttons', icons which can be activated with the mouse. Two of them are labelled respectively with forward and reverse arrows (⇦⇨), and these effect a move to the next card in either direction. The other icons, in the form of a circle or a square with words attached, initiate longer jumps: to the next or last heading, or to the exit from the program. Within some cards there are words or phrases highlighted in bold, and these also act as buttons to access definitions from the glossary.

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

LAC analysis: a guide

A way into a system for describing the material which a learner of EFL has met in class, with a view to using the resulting description as a basis for relevant assessment

© Andrew Harrison 1994

click here to continue

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Card system

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Card system

To follow the Guide you work with a set of cards, like this one. If an explanation fills more than one card, three dots appear in the bottom right hand corner. Clicking on the ⇨ button at the bottom of the card takes you to the next card; clicking on the ⇦ button takes you to the previous one. The 'last head' button allows you to skip back to the start of the current heading, and 'next head' to the start of the next one. You can leave the Guide at any time using the exit box at top left. ...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Headings

Card system ⇨
Introduction ⇨
Entries - reference ⇨
Entries - task ⇨
Entries - context ⇨
Entries - skills ⇨
Entries - language essentials ⇨
Entries - processing ⇨
Glossary - ⇨

You can return to this card at any time through 'exit' & then 'headings'

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Card system *(cont'd)*

There are also 'buttons' within the text boxes which lead to a Glossary of terms used in the Guide. These buttons are in bold type.

(Now try clicking on the word 'Glossary' above.)

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Introduction

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Introduction

Aims

The analysis is intended to record what learners have been presented with in class, and when enough information has been recorded, to act as a framework for the writing of assessment materials which will be relevant to what has been taught and/or learnt.

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Introduction *(cont'd)*

Assessment

The assessment materials will take the form of 'scenarios'. These are situations in which the learners can use what they have learnt for realistic purposes. The idea is that the results of the assessment will help learners to see how much they have achieved in the language and to encourage them to go on to make further progress.

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Introduction (cont'd)

Analysis sheets

The analysis is recorded on sheets listing elements which occur in any exchange of meaning. It treats the learning material as a series of tasks which the learner has to complete. You, as analyser, are to look at these tasks from the learner's point of view ('What do I have to do?'), with the help of defining questions which keep this perspective in mind.

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Introduction (cont'd)

Filling in the Analysis sheets

Cards with the title 'Entries' explain what you are to write under each sub-heading on the Analysis sheet. For Analysis sheet 1 there are three categories of entry: closed, closed +, and open. For Analysis sheet 2 entries take the form of numbers which indicate grades. (These will be explained when you get there.)

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Introduction (cont'd)

Entries

After identification of the coursebook and unit to be analysed, the Analysis sheets require a definition of each task in the first column, then entries in eighteen more columns, under sub-headings. The Analysis sheets show how the columns are laid out, with sheet 1 on one side of the paper and sheet 2 on the other. Look at them now, briefly, to get an idea of the system.

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Introduction (cont'd)

Link to assessment

The information which you build up on the Analysis sheets as you complete them will eventually represent a pattern of factors in the learner's completion of the various tasks. This pattern will become the framework for writing appropriate assessment materials.

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Introduction (cont'd)

What to write

The cards tell you what to write in each column. A set of possible choices follows after the instruction ENTER

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Introduction (cont'd)

KIt

Now check that you have available all that you will need to carry out the analysis, that is:
. sample of a completed Analysis sheet (both sides)
. blank Analysis sheets
. student's course book
. teacher's course book (if any)
. pencil, rubber
. ruler (useful, but not essential)

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Introduction (cont'd)

Working space

Your communication with the computer will continue to be through the mouse, so you will not need the computer keyboard for the moment. Put it out of the way and use the resulting space to lay out the course book(s) and blank Analysis sheets. You will also need to keep the sample Analysis sheet within reach.

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries - reference

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (1): reference

The first entry is a reference to the part of the course you are going to analyse.

Across top of Analysis sheet 1, in the long box labelled 'ref:'
ENTER name of course, unit, page etc...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries - task

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (2): task

You can now start on the first row of entries.

First of all, to count as a task, the activity must appear in the student's book (ie not in the teacher's book alone).
Secondly, there will be some form of rubric (find this term in the Glossary). The task starts with the rubric and ends when the instructions have been carried out. Remember: a preparatory rubric constitutes a task on its own. ...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (3): task *(cont'd)*

column: 1
heading: task

ENTER the title of the exercise as given in the course book, then under it a sequential reference letter a, b, c,... to indicate each task as you identify it.

(If you are now on your 2nd or subsequent row and want to get back to the last explanation card you saw, click here: Tasks 2+) ...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (4): task *(cont'd)*

column: 1
defining question: *what do I have to do with this item?*
list: closed+
ENTER one of these 10 options, or if none of them is appropriate, one of your own:
act out complete discuss do find
give own guess invent tell understand
followed by an indication (in brackets) of what it is that is to be acted out, completed... etc (see sample Analysis sheet).

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries - content

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (5): topic

column: 2
heading: topic
defining question: *what are we talking about?*
list: [open]
ENTER short description of topic (2-3 words)
examples: TV, sport, requests/offers, adjectives, verb forms.

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (6): situation

column: 3
heading: situation
1st entry: who
defining question: *who is involved in this exchange?*
list: closed
ENTER two of these 8 options:
self author course ch/s L1 person
originator partner/s stu T

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (7): situation *(cont'd)*

column: 3
2nd entry: where
defining question: *where am I when this exchange takes place? or is the physical context irrelevant to this task?*
list: closed and open
ENTER one of these 3 options ('topic place' is open, so enter one or two words of your own):
classroom [topic place] neutral

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (8): set

column: 4
heading: set
defining question: *who else is involved with me in this exchange?*
list: closed
ENTER one of these 4 options:
lone pair group class

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (9): culture

column: 5
heading: culture
defining question: *what do I perceive as the background culture to this exchange?*
list: closed +
ENTER one of these 6 options, or if none of them is appropriate, one of your own:
own local CHE British WEuropean worldwide

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries - skills

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (10): receptive

column: 6
heading: receptive
defining question: *how does this input reach me?*
list: closed
ENTER one or more of these 4 options:
L R vis b/board

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (11): productive

column: 7
heading: productive
defining question: *what kind of output do I have to create in order to carry out this task?*
list: closed
ENTER one or more of these 6 options:
S W act draw tick none

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries - language essentials

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (12): structure

column: 8
heading: structure
defining question: *what grammatical structure(s) do I have to use to carry out this task?*
list: [open]
ENTER structure(s) which are essential for my completion of this task, ie the minimum low without which it would be impossible for me to do it.

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (13): lexis

column: 9
heading: lexis
1st entry: field
defining question: *what is the semantic field within which this task is operating?*
list: [open]
ENTER a 1- to 3-word description
examples: fitness, a bank loan, nationalities

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (14) lexis *(cont'd)*

column: 9
2nd entry: vocabulary
defining question: *what words are essential for my completion of this task?*
ENTER sample of the minimum few words without which it would be impossible for me to do this task.

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (15) lexis *(cont'd)*

column 9
3rd entry: function
defining question: *what can I do with these words?*
list: open
ENTER the use to which I put words for this task
examples:
discussing, narrating, giving information, making comparisons, finding differences...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (16): phonology

column: 10
heading: phonology
defining question: *what sounds do I have to make to carry out this task?*
list: [open]
ENTER only those phonetic features which are specifically mentioned in the coursebook as learning points in this task.

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (17): discourse

column: 11
heading: discourse
1st entry: relationship
defining question: *what is my relationship with other participants in this task?*
list: closed
ENTER one of these 4 options:
equal sub sup outsider

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (18): discourse *(cont'd)*

2nd entry: assumptions
defining question: *what needs to be common knowledge between me and the other participants to the extent that we do not have to refer to it?*
list: [open]
ENTER short description of area
example: conventions at an auction (students hear excerpts from a village auction ... & must try to write down how much each person paid CEC2 p88); form filling (HI p5); classroom conventions ...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (19): discourse *(cont'd)*

3rd entry: expectations
defining question: *how high are the expectations of my understanding beyond the literal text?*
ENTER a rating, low 1 to high 3
examples:
low - Write a description of someone you know... parts of the body, clothes, ... HI p 13
high - Listen ... Dick is talking to 3 different people... (work out from what they say who and where they are) HI p 55

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries - processing

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (20)' processing

Use Analysis sheet 2 for entries under the heading 'processing'. Entries here take the form of grades (rather than descriptions, as in Analysis sheet 1). There are 8 sub-headings, representing elements in processing: input, response, load, search, link, know, individual, demand. For each element there are: . a defining question;
. keywords for possible aspects of the element;
. examples of different levels of operation.
...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (21) processing *(cont'd)*

You are to grade the task by comparing what you find in the course book with the examples. The first card for each element gives the extreme grades, top and bottom, to help you judge what grade to enter on the Analysis sheet. The next card gives examples of the other grades, if needed. The figure at the top of each column (eg /5) is a reminder of the top grade. When you have entered your grade, add a short comment (like those in the examples) to help indicate why you gave that grade.
...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (22): processing *(cont'd)*

First of all, to identify the task and link this sheet back to Analysis sheet 1:

ENTER a repeat of the reference letter (a,b,c...) and the task as you described it in column 1.

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (23)' processing *(cont'd)*

sub-heading: 12 input
defining question: *how dense is the input and how difficult is it to take in?*
keywords: density, quality
ENTER a rating, maximum 8
examples:
high 8: understand joke, R TV listings & discuss, R authentic tourist text, R re politician, L L1 discussion
low 1: L sounds for stories, vis pics for 'going to', R model sentences, R headlines
...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (24): processing *(cont'd)*

12 input - more examples:

5: L interview ques (role play), L sentences with elisions (& restore), R business ad & letters
4: R match functions & sentences, review exs & find rule, L sts' views on views of R text, L group discn
3: L dialogue, R answer (& find question), R role play rubric & L partner doing it, vis R 2 postcards
2: L partner, R story title (& guess), R form
...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (25): processing *(cont'd)*

sub-heading: 13 response
defining question: *how quick/psychologically prepared a response is expected from me?*
keywords: pressure, speed, competition
ENTER a rating, maximum 5
examples:
high 5: L numbers & W, speedread & S to class, competitive (game, etc), L+draw on map
low 1: in own time, homework, W after text model, study+tick, no competition intended
...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (26)' processing *(cont'd)*

13 response - more examples:

4: S to class, report to T & class, S if T chooses
3: S to group, S to class if choose to, competition optional
2: S to partner, L & find words in transcript
...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (27)' processing *(cont'd)*

sub-heading: 14 load
defining question: *how many inputs are there and how many are happening at once?*
keywords: quantity, simultaneous, N skills
ENTER a rating, maximum 7
examples:
high 7: 30 lines of transcript, R 700 words, all class involved, L+R summary, review whole unit
low 1: L sounds, R grammar point, L partner, R vis & 3 ques
...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (28)' processing *(cont'd)*

14 load - more examples

6: R news & 2 grammar ques & L discn, chatred class discn, R half page of text & vis
5: 8 verbs, 4 lines of text, 15 short lines of dialogue
4: R 8 sentences & functions & L partner's view
3: 2 course chs & pair, group discn, LR 4 lines of text, R 2 chs & T explanation
2: own & partner's ideas, 10 wds & dictionary definitions

last head next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (29)' processing *(cont'd)*

sub-heading: 15 search
defining question: *how much imagination/logic/ memory do I need to carry out this task?*
keywords: find, invent, work out, remember
ENTER a rating, maximum 6
examples:
high 6: improvise, formulate narrative, work out pattern from examples, see implications
low 1: fit to pattern, produce in parallel, remember own experience ...

last head next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (30)' processing *(cont'd)*

15 search - more examples:

5: find own ideas on topic, work out what functions used, find/remember jokes about topic,
4: find function in text, remember L & compare, work out connections in text, find info in L text
3: find words in R text, invent explanation for partner, remember models, find elisions, recall own predictions
2: recall earlier in unit, remember weather ...

last head next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (31)' processing *(cont'd)*

sub-heading: 16 link
defining question: *how difficult is it for me to transfer word to idea, text to meaning, or vice versa?*
keywords: apply, extract, compare, rearrange
ENTER a rating, maximum 6
examples:
high 6: extract discussion & S, apply principle to this case, express ideas on ethic, explain socio issue
low 1: apply greetings to situation, apply learnt rule, compare pics/text, explain weather forecast ...

last head next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (32)' processing *(cont'd)*

16 link - more examples

5: apply words to sounds, extract news, compare grammar/meaning, rearrange to show implications
4: apply phrases to situations, link+Q&A, extract ch's views, make predictions, rearrange reported speech
3: compare R&L, apply in role play, compare egs & grammatical form, rearrange for summary
2: extract types from list, compare self with form to fill, extract points of R text ...

last head next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (33): processing *(cont'd)*

sub-heading: 17 know
defining question: *how far is the task/content/ situation likely to be familiar to me?*
keywords: know, recognise, connect
ENTER a rating, maximum 6
examples:
high 6: specialist topic, principles of organisation, reported speech system
low 1: numbers, own view, own culture, info about self, done this just now ...

last head next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (34)' processing *(cont'd)*

17 know - more examples

5: stereotypes, about nationalities, recognise grammar pattern, know social conventions
4: connect with situations, know restaurant system, guess what comes next, know way round dictionary
3: connect own/others' foods, about descriptions of people, about holidays/travel, all unit to summary
2: done this earlier, about family life, reread text, everyday events, own past holiday ...

last head next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (35): processing *(cont'd)*

sub-heading: 18 individual
defining question: *how committed am I likely to be to the task, how relevant is the content to me?*
keywords: own, interest, useful, intriguing
ENTER a rating, maximum 3
examples: (reminder: high = difficult)
high 3: hypothetical, no necessary identification, no relevance, no commitment, why bother
low 1: own convictions, own experience, facts about self, own views, unusual task ...

last head next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (36): processing *(cont'd)*

18 individual - more examples

2: if want, own predictions, could be fun, could be useful, here we go again

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (37): processing *(cont'd)*

sub-heading: 18 demand
defining question: *what is expected in my response?*
keywords: fluent/accurate, vague/exact, close/open
ENTER a rating, maximum 8
examples:
high 8: accuracy essential, fluent & accurate, L & guess what, make dialogues to fit, got to be plausible
low 1: fluency not expected, underline grammer examples in text, discuss if want, joint effort, match 2 givens

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (38): processing *(cont'd)*

18 demand - more examples

5: reasonably accurate, S cogent arguments, find Q forms for ans, tell joke, W summary R text
4: accuracy helps, prompts to narrative, substitution, known verbs applied, closed format + closed content
3: fluent but not necessarily accurate, limited by Q, draw & discuss, open wording + closed facts
2: discuss with partner, match functions & speech, open wording + own facts

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (38): what next?

You have now reached the end of the first row of entries on the Analysis sheets.

What an achievement!

Now you need to go back to the first column and start to record the analysis for the next task.
But the rules for deciding what constitutes a task are a little different for the second and subsequent rows of entries: the next card explains.

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (40): tasks 2+

To define what is to be considered a task for the second and subsequent rows of entries on the Analysis sheet, make sure it is in the student's book, as before, and then decide whether the rubric is explanatory (part of the task) or preparatory, which means that it is counted as a task on its own.

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Entries (41): tasks 2+ *(cont'd)*

Then ask in addition:
. Is it a new topic? (col 2)
. Do both 'who' and 'where' change? (col 3)
. Does 'bet' change? (col 4)
A 'yes' to any one of these three questions, taken in order, indicates that the activity is to be entered as a new task. If 'no' to all three of them, not.

The right-hand arrow below now takes you back to the entry for column 1, row 2.

...

last head

next head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

exit

Glossary

last head

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

closed

Closed lists are used for entries in Analysis sheet 1 when there is a finite set which covers all alternatives, for example under 'receptive' (column 6), Listen, Read, Visual, Manual. No other entries are allowed unless you find (exceptionally) that you need an additional category. If so, enter it, but in CAPITALS so as to draw attention to it as an exception.

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

rubric

Instructions in the course book explaining how to carry out a task.
An explanatory rubric tells you how the task is to be done, and is an integral part of that task (eg Read the story...);
A preparatory rubric introduces topics, structures, vocabulary, etc, and will constitute a task on its own (eg Which of the following can you find in the picture?).

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

open

Open lines are indicated by [square brackets]. They are used for entries in Analysis sheet 1 when the information required is specific to the unit being analysed, so you have to insert one or two words of your own to answer the defining question.

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

closed +

Closed + lines are used for entries in Analysis sheet 1 when the set given may not cover all alternatives, for example under 'task' (column 1), where 10 possible activities are listed, but you may need to add others.

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

self

learner as participant in exchange

examples:
author - self, self - partner

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

author

writer of course book as the other participant in task with learner, mainly in mechanical exercise which has no social content
example:
author - self Look at the example. Then write the past tenses and past participles of the verbs below... know steal go drink... CEC2 p 52

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

ENTER

This is the signal for you to choose what alternatives to write on the Analysis sheet in the columns indicated, and then to write it in.
All alternatives are defined in the Glossary, which is set out on cards in the same format as this present one.

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

Glossary

The Glossary gives an explanation for every term which appears in bold type in the Guide, and clicking on the word leads to the explanation.

To return to the Guide when you left it, click on the 'Go back' button below.

Go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

grades

Entries on analysis sheet 2 consist of numerical grades. There are eight elements which are to be graded according to likely difficulty for the learner (easy = 1, difficult = 7). A survey of several course books has suggested that the range for different elements varies from 1 - 7 to 1 - 3. The examples given for each element have been drawn mainly from these course books.

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

processing

The mental processes required of a learner in coping with the task, or in terms of a defining question:
what kind of mental ambidexterity do I have to undertake to carry out this task?

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

cut out

the task is to take part in any form of role play, including practice of a structure or function
example:
Make up similar dialogues. Student A You have a room in Student B's house. Student B you are a landlord/ landlady...Student A wants to borrow B's newspaper/turn the television on... HI p55

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

complete

the task is to fill gaps in a text
example:
Look back at the text and decide whether 'will' or 'may' goes in each blank. Not all banks in small towns in Britain _ change foreign money for you; If you stay in a hotel in Britain, it _ be expensive. CEC2 p24

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

discuss

the task is to discuss topic/exercise with others (pair, group, students in class as a whole)
example:
How are you both... with a partner, compare them... HI p64

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

do

the task is to carry out instructions for grammatical or other mechanical exercise which has no social context
example:
Look at the sentences... Does sentence (a) refer to the present or the future? HI p65

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

find

the task is to find information, structure, words...
example:
Play the memory and ask students to note the four expressions they think are exactly the same in the box and in the conversation CEC p107

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

give own

the task is to give own view, argument, experience...
example:
What kind of person wrote this article? What makes you think so? HI p37; What are the most common causes of accidents in your country? Colvald p51

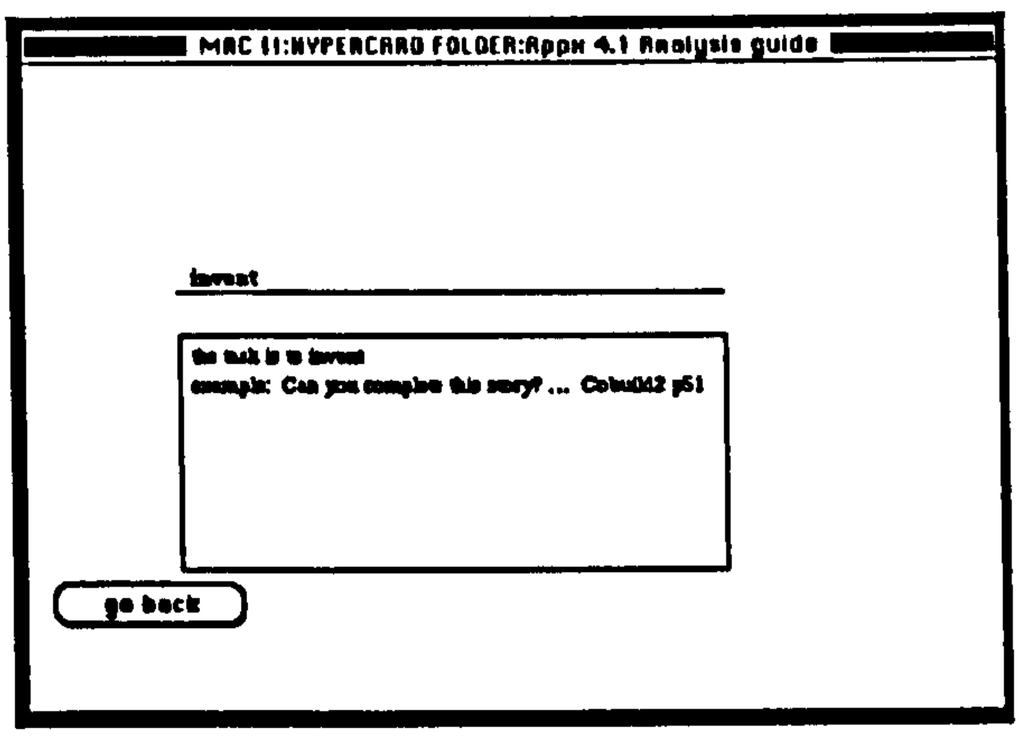
go back

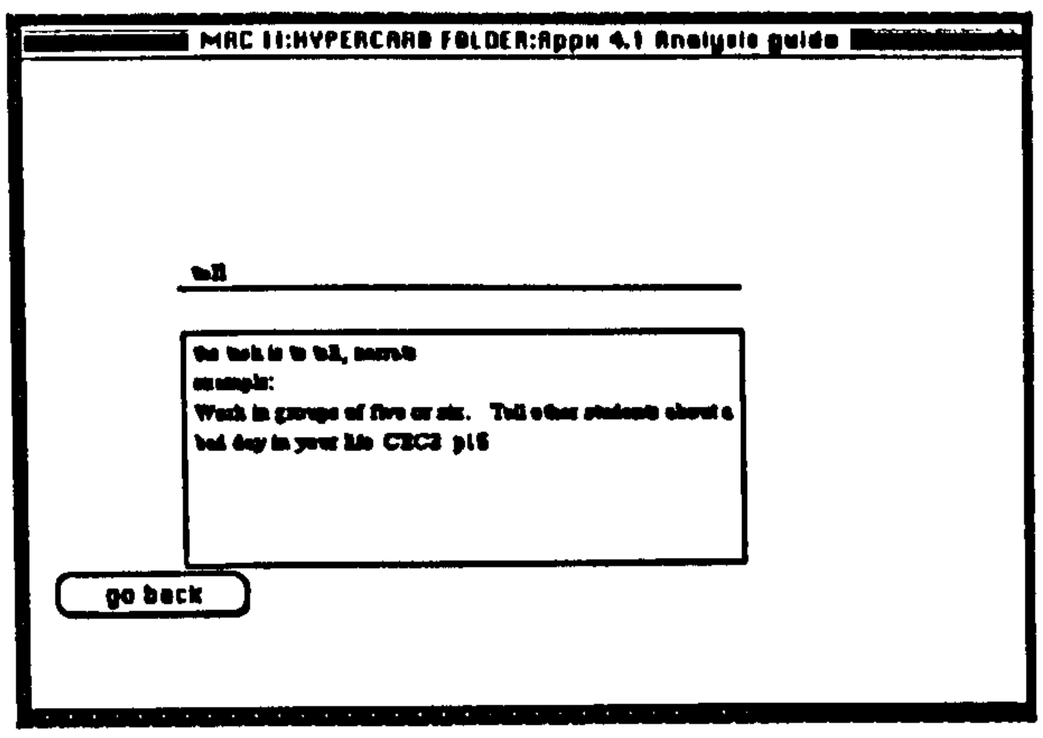
MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

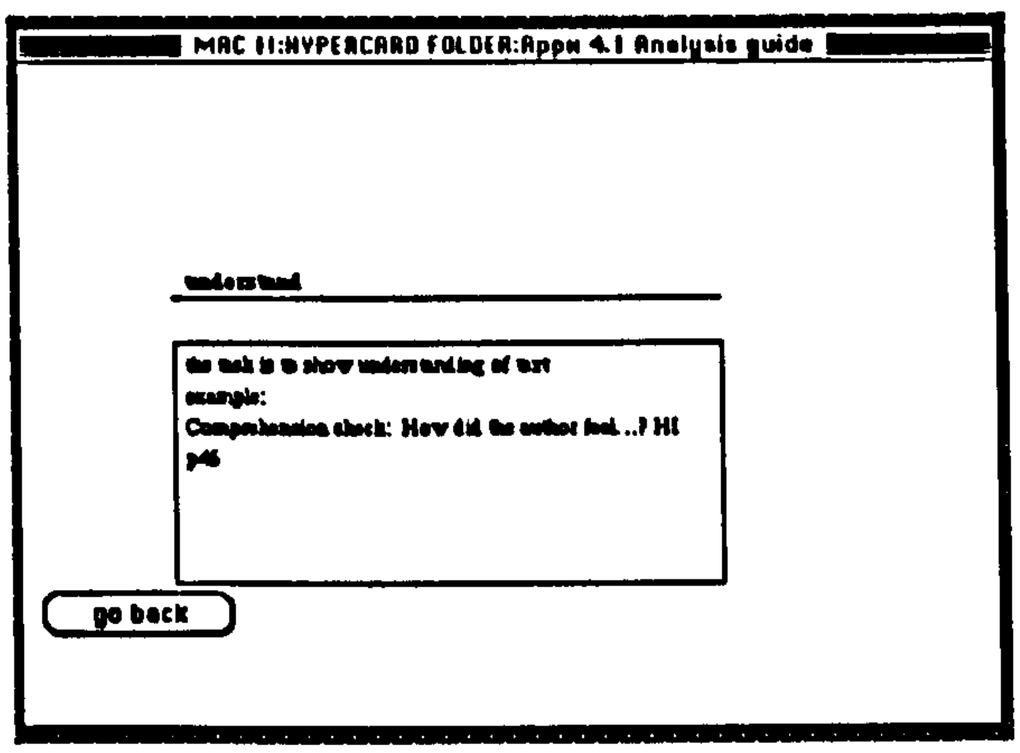
guess

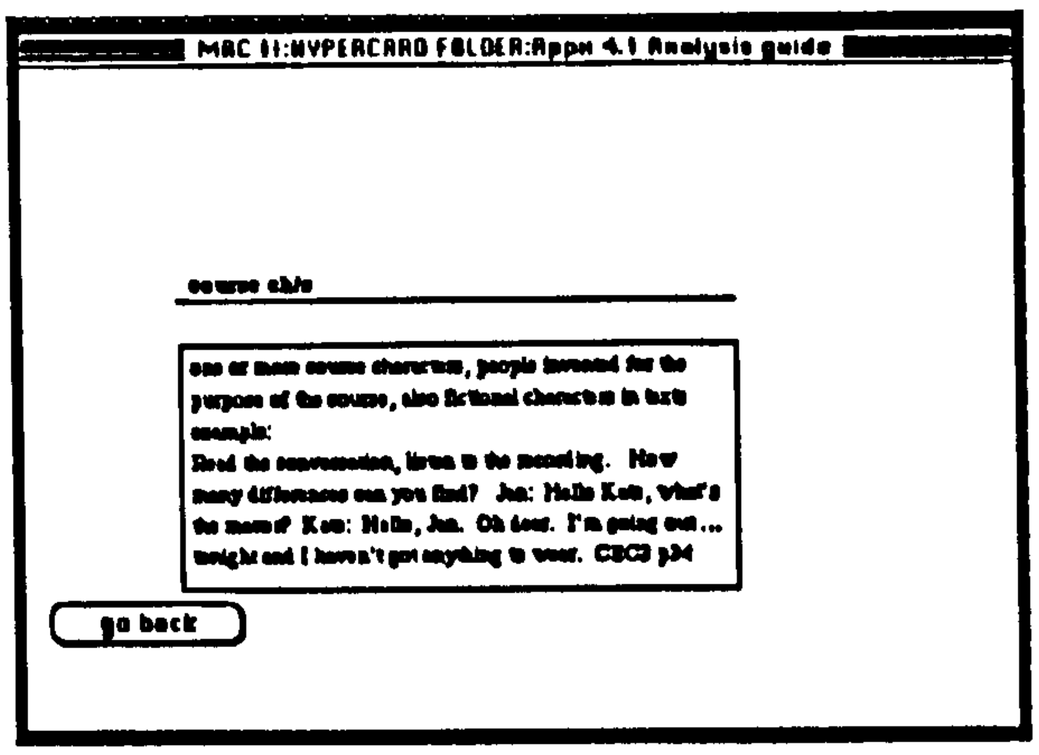
the task is to guess!
example:
Think of a country and say 4 sentences... students try to guess CEC2 p51T

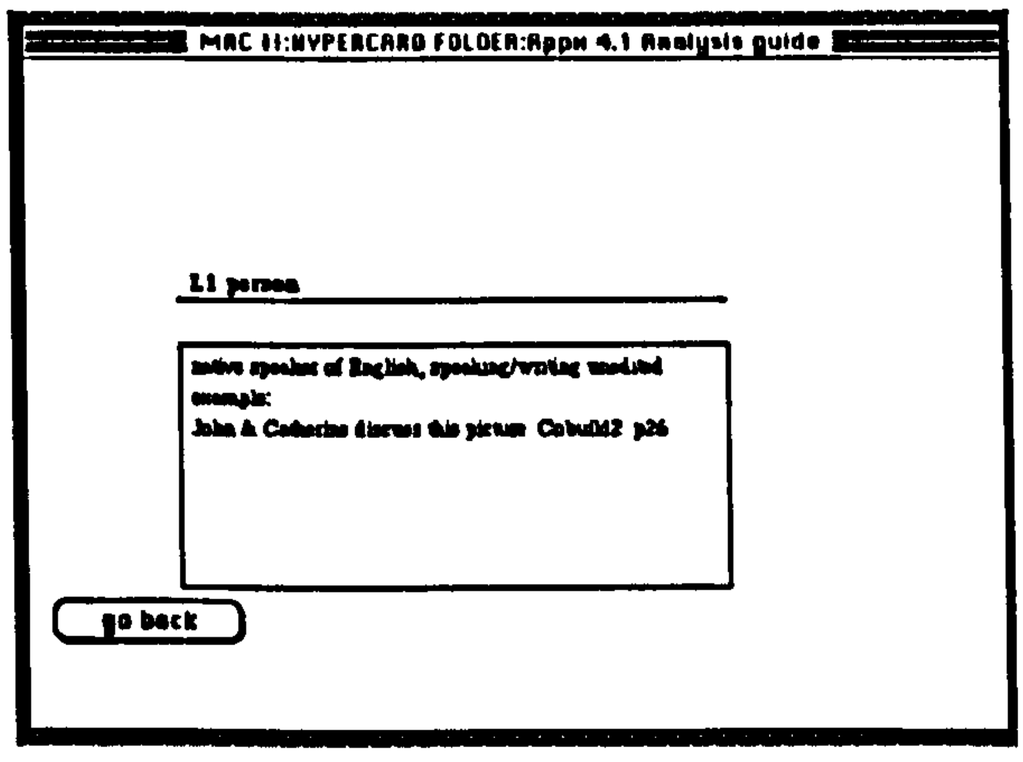
go back

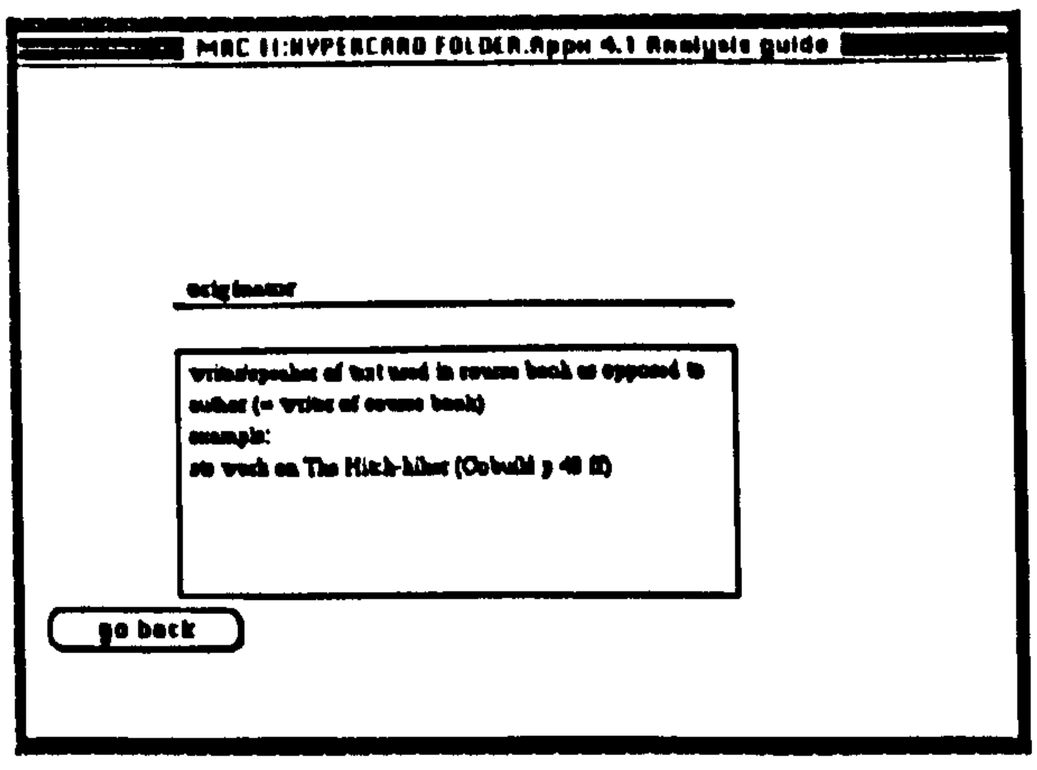

MAC II:HYPERCARD FOLDER:Appn 4.1 Analysis guide
invent
the task is to invent
example: Can you complete this story? ... Cobuild2 p51
go back

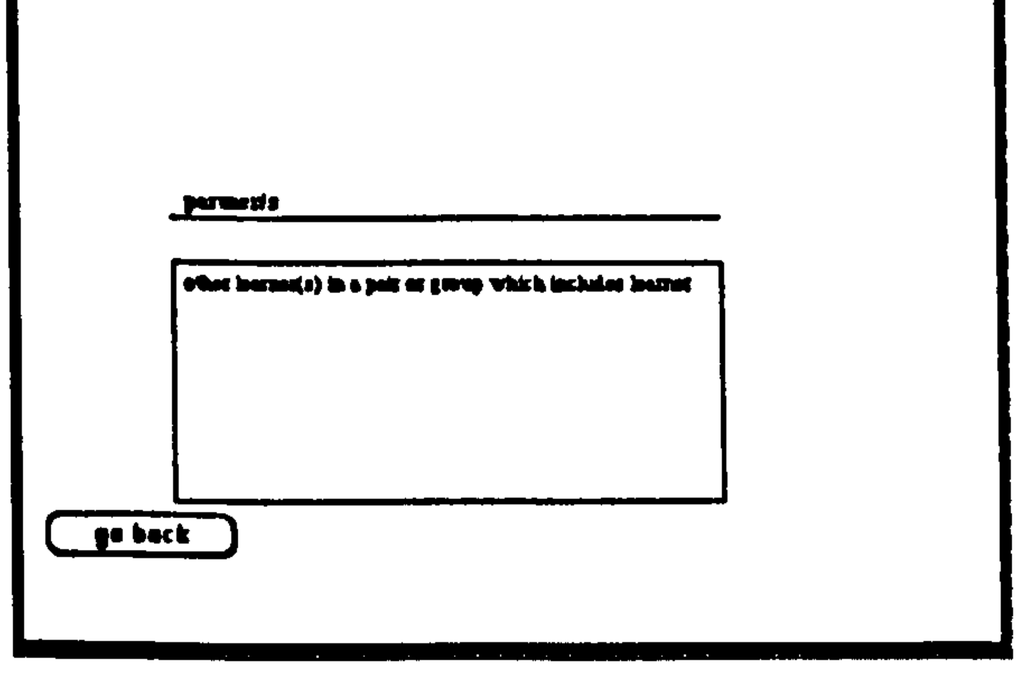

MAC II:HYPERCARD FOLDER:Appn 4.1 Analysis guide
tell
go back

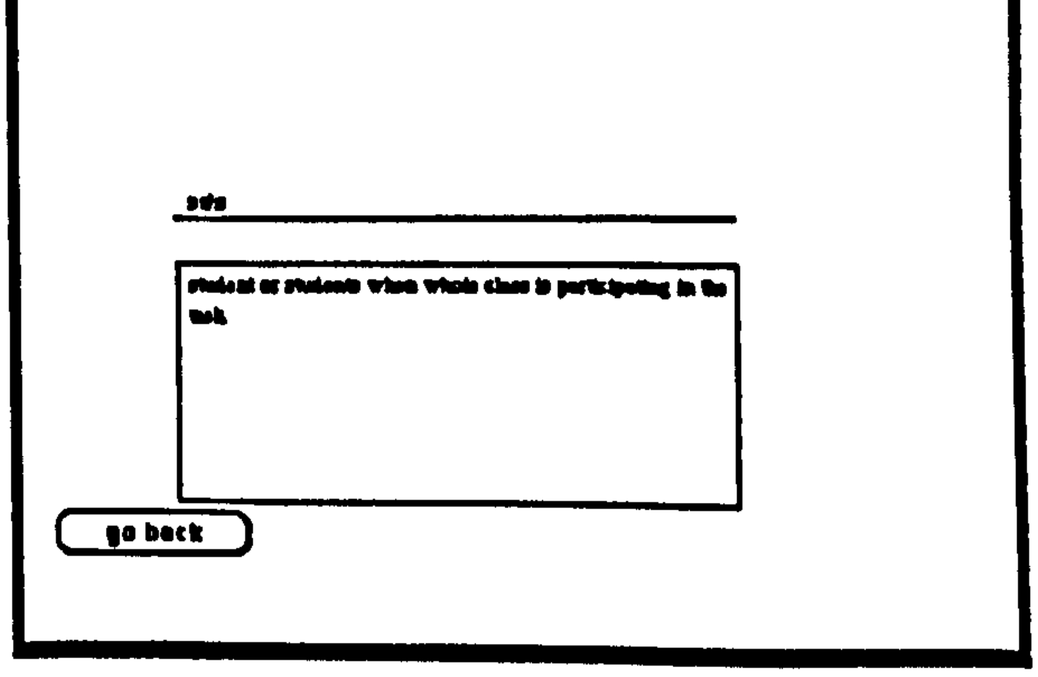

MAC II:HYPERCARD FOLDER:Appn 4.1 Analysis guide
understand
the task is to show understanding of txt
example:
Comprehension check: How did the author feel ..? Hi p46
go back


MAC II:HYPERCARD FOLDER:Appn 4.1 Analysis guide
course chfs
one or more course characters, people invented for the purpose of the course, also fictional characters in txts
example:
Read the conversation, listen to the recording. How many differences can you find? Jan: Hello Ken, what's the matter? Ken: Hello, Jan. Oh dear. I'm going out ... tonight and I haven't got anything to wear. CEC3 p34
go back


MAC II:HYPERCARD FOLDER:Appn 4.1 Analysis guide
L1 person
native speaker of English, speaking/writing modern
example:
John & Catherine discuss this picture Cobuild2 p26
go back


MAC II:HYPERCARD FOLDER:Appn 4.1 Analysis guide
originator
writer/speaker of txt used in course book as opposed to author (= writer of course book)
example:
as work on The Hitch-hiker (Cobuild p 48 ff)
go back


MAC II:HYPERCARD FOLDER:Appn 4.1 Analysis guide
partner/s
other learner(s) in a pair or group which includes learner
go back


MAC II:HYPERCARD FOLDER:Appn 4.1 Analysis guide
s/s
student or students when whole class is participating in the task
go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

T

teacher as participant in exchange
example:
T - st: dictation EFSA p767; do the comprehension questions HI T9 p6

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

classroom

place where talk occurs if not [topic place] or neutral, is some of general discussions about student's own experience and opinions (social group being T and st) or of exercises without social dimension (learn grammar, learn), ex

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

[topic place]

description of place where exchange occurs, if not classroom or neutral. In 'quotes' if let's pretend, eg role plays.
example:
section: Students hear excerpts from a village auction CEC2 p69T; 'hotel' You are in a hotel. One of you is the receptionist, the other a guest HI p18

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

neutral

when place of exchange is irrelevant to topic, as when discussing last night's TV in the queue at the post office, as opposed to buying stamps at the counter.
example:
A: Where are you going for your holidays this year? B: We haven't decided yet HI p66

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

learner

when learner is working on own

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

pair

when one other learner participates in task with learner

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

group

more than pair but fewer than whole class

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

class

the whole class is involved together in doing the task

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

own

culture of learner's own background
example:
Students talk [in groups] about their local shops in their home town or village Cotrell&2 p60

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

local

culture local to classroom or to all learners
example:
Discuss products of own country (EFSA p 13)

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

CHE

common human experience of a personal kind, as against 'world wide', which is cultural
example:
Do you get up as soon as you wake up? Do you have breakfast before or after you get dressed? CEC2 p70

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

British

culture within which text or task operates is British, as opposed to local or W European

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

WEuropean

culture within which text or task operates is West European, as opposed to local or British, and when (particularly) to include US and white commonwealth

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

British

culture within which text or task operates is British, as opposed to local or W European

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

L

listening as a receptive activity

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

R

reading as a receptive activity

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

vis

visual representation as part of input, to include illustrations in course book (but only when essential to task, not when just decoration), graphs, charts etc, whether on paper or OHP, etc

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

b'board

black/white board or other writing by teacher for whole class, as opposed to visual representation

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

s

speaking as a productive activity

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

w

writing as a productive activity

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

act

carry out some action in response, usually as evidence of understanding

go back

MAC II:HYPERCARD FOLDER Appx 4.1 Analysis guide

d:nv

use some non-writing means to express meaning (counterpart to visual in column 1)

go back

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

tick

marking non-verbally to indicate a choice (a productive activity) in tasks requiring eg circling of a multiple choice option, choosing T or F to indicate true/false, numbering, lettering, underlining...etc

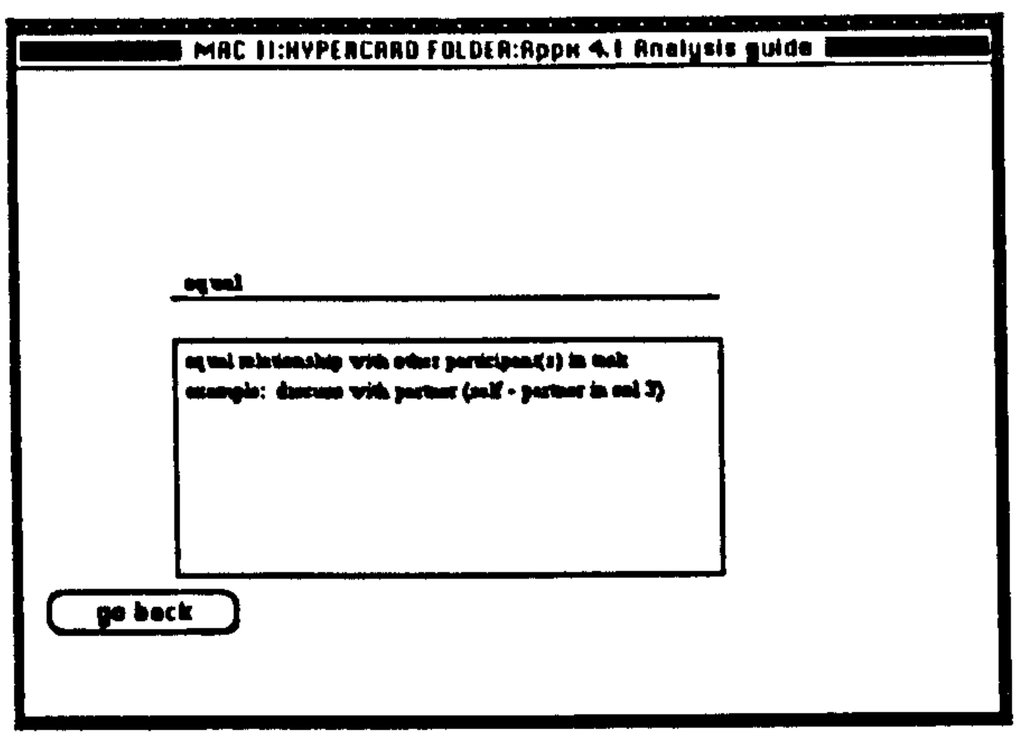
go back

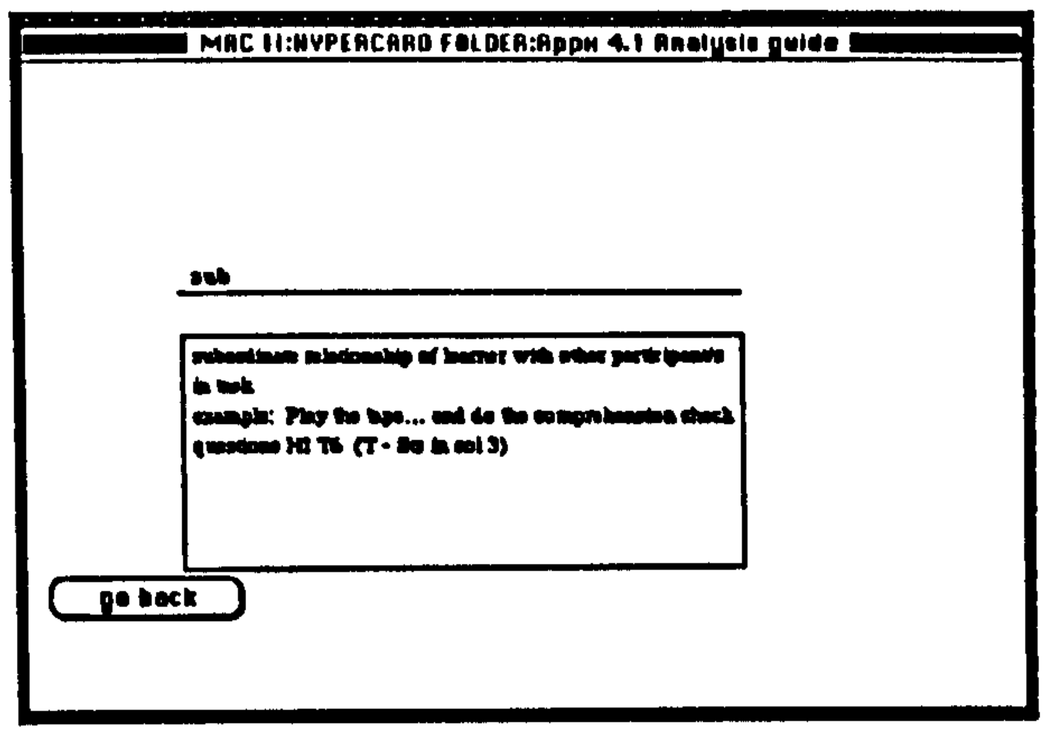
MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide

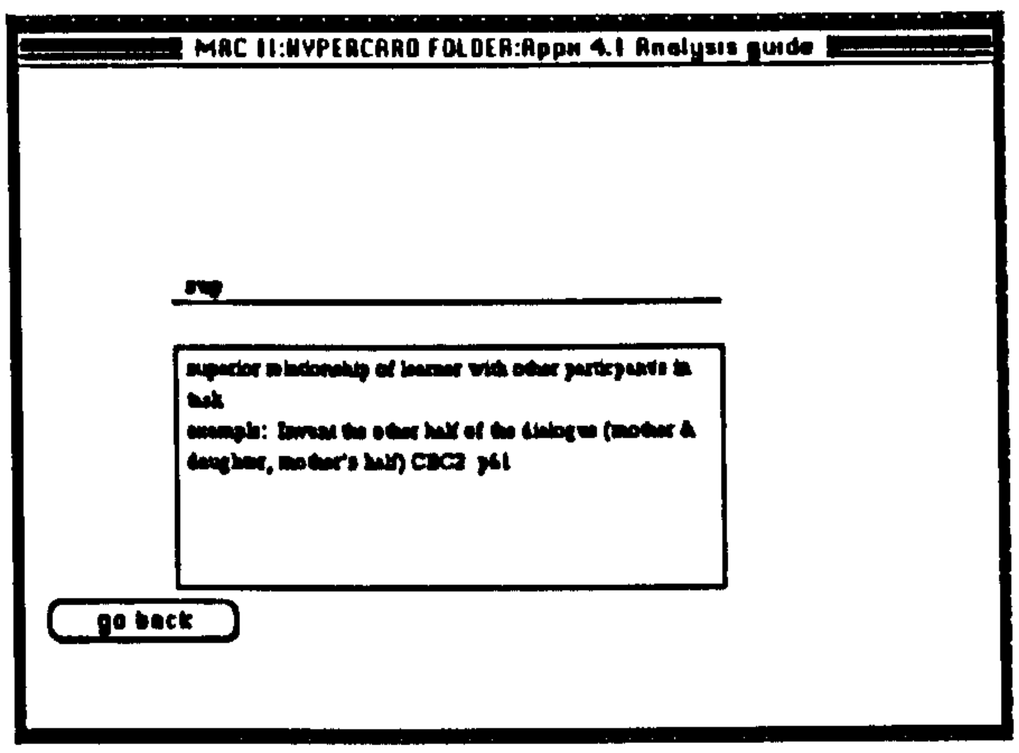
note

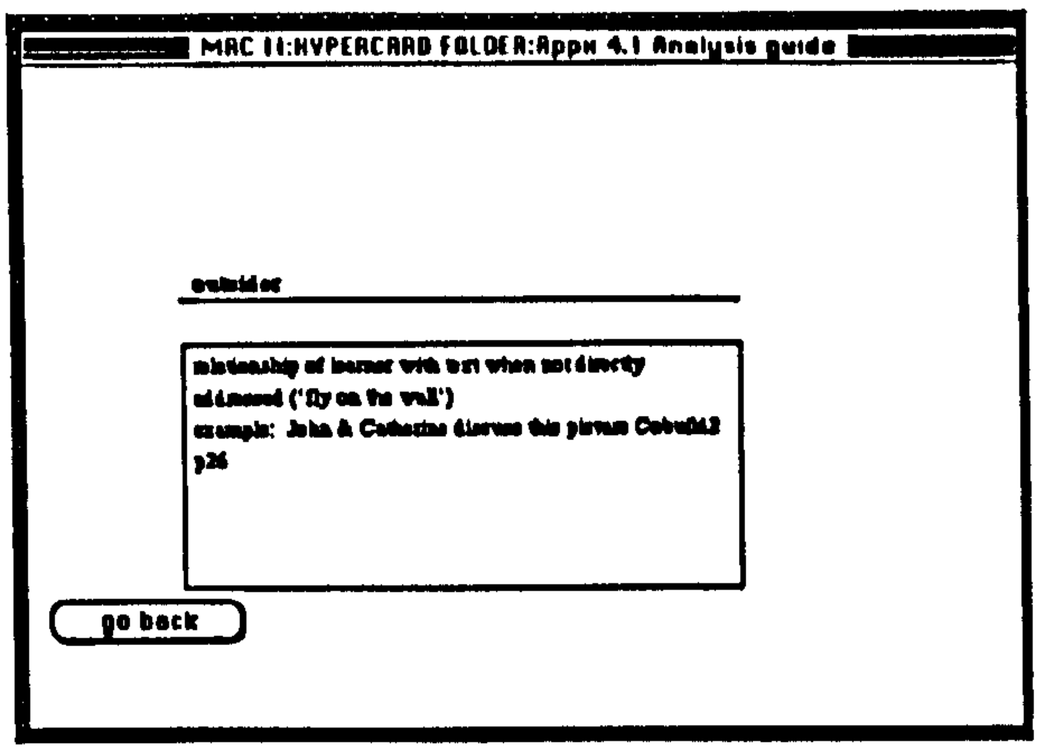
when an entry can be made under a sub-heading

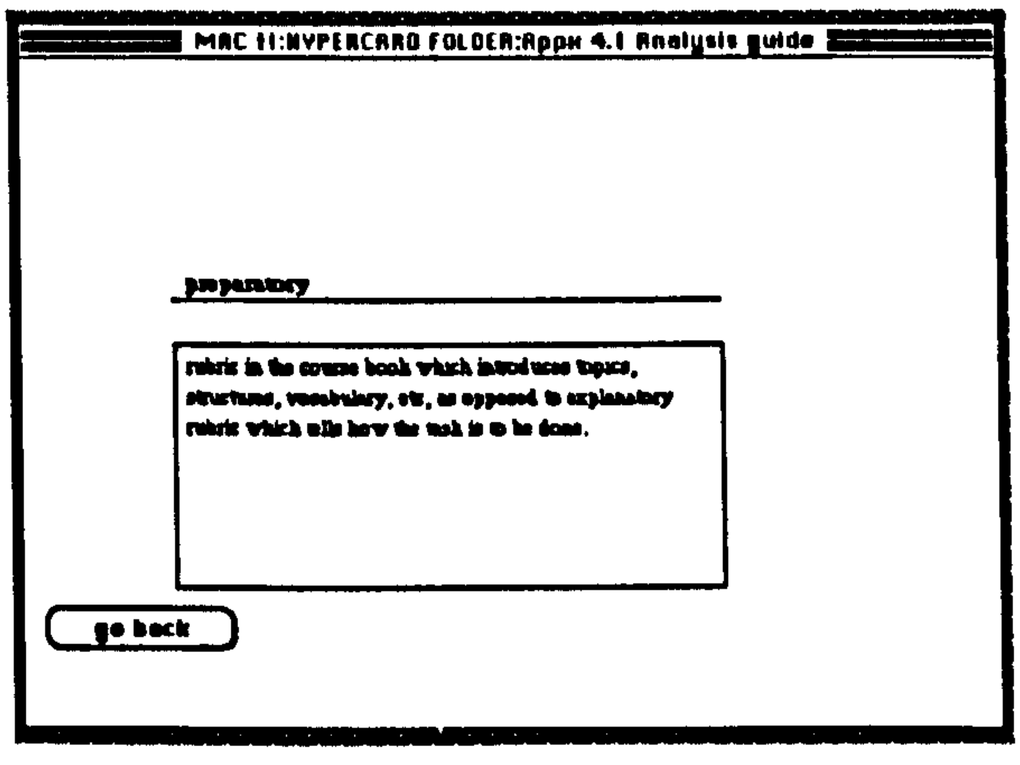
go back

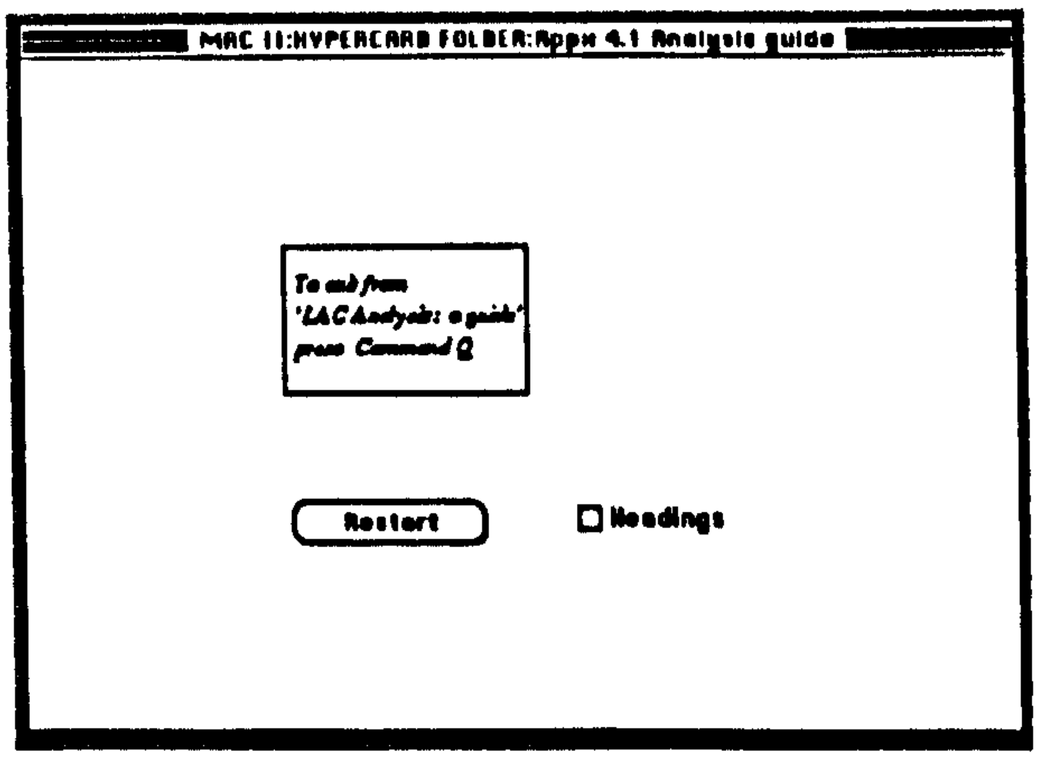

MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide
equal
equal relationship with other participant(s) in task
example: discuss with partner (self - partner in col 3)
go back


MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide
sub
subordinate relationship of learner with other participants in task
example: Play the tape... and do the comprehension check questions NI T6 (T - Su in col 3)
go back


MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide
sup
superior relationship of learner with other participants in task
example: Invent the other half of the dialogue (mother & daughter, mother's half) CBC2 p61
go back


MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide
outsider
relationship of learner with art when not directly addressed ('fly on the wall')
example: John & Catherine discuss this picture Cobuild2 p26
go back


MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide
preparatory
rubric in the course book which introduces topics, structures, vocabulary, etc, as opposed to explanatory rubric which tells how the task is to be done.
go back


MAC II:HYPERCARD FOLDER:Appx 4.1 Analysis guide
To exit from 'LAC Analysis: a guide' press Command Q
Restart
Headings

Appendices

Appendix 5.1

Course book analysis

This appendix consists of reproductions of students book, teachers book and recordings from *Cobuild* 2, and the relevant analysis sheets as a demonstration of the analysis system in action.

Appendices

Appendix 5.1.IS

12 Languages you've learnt

a What languages can you speak? How many foreign languages have you learnt, or tried to learn?

12b **b** Catherine talked to someone called Stephen about the languages they knew. Listen and make a list of the languages they knew, and say how good they were at each.

12c **c** Many people say that the British are very lazy about learning foreign languages, because they think that when they go abroad they will usually be able to find someone who speaks English.
To find out if this was true, we asked a variety of British people from different walks of life about their language learning experiences.
Some of their comments and stories are written here. Read them. Do you think the British *are* lazy about learning languages?

d Which of these people do you think is most serious about learning languages?
Which of them do you think has the best sense of humour?

e What do you think is the best way of learning a foreign language?

By reading a lot	Reading a dictionary
Doing grammar exercises	Translating
Trying to speak it, even though you make mistakes	Watching English or American films on TV

f What experience do you have of learning English? Write a short paragraph.

g Language survey

Design a survey form. Practise the questions you might ask. Ask other people in your class.

Tell each other which person you think speaks the most languages.
Who has learnt the most languages in the class as a whole?

SB

Appendices

Appendix 5.1.IT

12 Languages you've learnt

Aims: 1 To encourage students to read for interest and relevant details only.
2 To give students an opportunity to react personally to other people's ideas and to relate them to their own feelings and experiences.

Lexis: **apart from** [L], **(any) degree**, **foreign** [L], **learning**, **literature**, **not**, **particular**, **particularly**, **Russian**, **Spanish**, **struggle** [L], **study**, **tiny (bit)**
Revision: **any more** (any longer), **get by in** (just manage), **(fairly) similar**, **that's about it**

SB12a PB12a

1 Don't go into these questions in detail (omit them altogether if you want to do 12c below). Instead tell students about your own language learning experiences. Say what languages you can speak, read and write now.

Listening SB12b 12b* **PB12**

2 Students listen to the tape and take notes about Catherine and Stephen. Suggest they make a table to fill in and give them a framework like the following.

Key:

	CATH.	STEPH.
FRENCH	learnt at school	can get by
GERMAN	learnt at school	no
SPANISH	no	no
ITALIAN	tiny bit	reads a little
RUSSIAN	no	no
LATIN	no	no
OTHER: MALAY, THAI JAPANESE	a tiny bit of Greek	American

When students have done 12c they can extend this table to include M. West, R. Turner, C. Egerton and E. Turner.

Listening SB12c 12c

3 Students read the introduction to 12c. Ask them if they feel that English people are lazy about learning other foreign languages. How many British people do they know who can speak their language really well/quite well/just a bit? Encourage them to comment if they wish.

4 Students read and then listen to reports 1–4, which are written by British people. Ask students for their reactions to these before they work in pairs to complete the table from 12b.

5 Pairs compare answers. An overhead transparency would provide a good focus for a further class discussion of answers.

Key: M. West: speaks some French and a bit of Russian
R. Turner: learnt French at school, was taught Latin and German but didn't learn either.
C. Egerton: learnt French and Spanish at school, learnt German at school and university and speaks Malay and Thai
E. Turner: studied French, and Latin at school.
All are native speakers of English

SB12d

6 Students will have already discussed 12b and c in some detail, so go over these questions to summarise.
Class vote: does the class agree that the British are really lazy about learning other people's languages? Let them vote: yes/no/just a bit.

SB12e

7 Students choose the 3 best ways of learning a language individually. Then put the class responses on the board and discuss. Ask students why they have selected their responses and what other ideas they have.

8 Tell students to find some statements in the text that apply to them personally. (Give students a minute to prepare their answers to this.) Then ask them: What are some of the most useful English phrases you know already? When might you/do you have to use English in your own country? Why are you learning English now?

Writing homework SB12f

9 If you have a lower ability class, students might find this framework useful as preparation for the paragraph. See also review page 13c.

Many/most/some people say that . .	the British never speak . . .
	the British are quite willing to learn . . .
	is fairly easy to learn.
	is quite difficult to learn.

We were taught . . . at school, but . .

I took a . . course some/many/two years ago, and/but . . .

I tried to learn . . . but I didn't get very far/do very well/remember very much.

When I try to speak . . ., it's always/often a . . . word that comes into my head instead.

My . . . has been particularly/quite/very useful.

Planning SB12g

10 Students' survey questions could be based on the conversation between Catherine and John. Students need to find out how many languages the others know and which they are the most fluent in. Encourage them to ask specific questions about skills: listening, speaking, reading, writing; often language learners will be very fluent in some skills and not in others.

11 Students write down some of the questions they will need to ask. Teacher provides input and assistance as needed. If students need more help, refer them to 13a on the review page, but warn them that they must choose questions that make sense, e.g. 'Where did you do?' does not make sense.

12 Students take turns to interview the other members of their groups. Ask them to appoint a group secretary to record results for the group. Help students collate their results simply, e.g. 'Everybody in this group spoke at least two languages, . . . and . . . . One person also speaks . . . We can all read and write . . ., and two of us can . . . .' and so on.

Report

13 Students either read the results of their surveys to the class or pass around a written report. Written reports could be collected later and put on the wall. The other students read or listen to find out who in the class speaks/has learnt the most languages.

Appendix 5.1.IR

12b

SB: How many languages do you speak?
CM: Erm, well, I learnt, erm English, French and German at school, I know a tiny bit of Italian and a tiny bit of Greek, from travels, but I can't say that I really speak any of the foreign languages to any degree, any more.
SB: Well apart from English and American which is fairly similar, I can get by in French and I can struggle through reading Italian and that's about it.

12c

Meriel West, Travel Agent, currently working in New York.
We were taught French at school, but very badly. We had to memorise lists of verbs, instead of being given sentences to learn which we could use in everyday speaking. I took another French course a few years ago, and we learnt some quite useful things. I also tried to learn Russian, but I did not get very far, although it was interesting. In fact I found a number of words have nearly the same pronunciation in both Russian and English.

Richard H. Turner, student of Engineering at Loughborough University.
The foreign languages I was taught at school were Latin, German and French. However, the only language I actually learnt was French.
(Although I 'picked up' English at the early age of one.)
My French has been particularly useful. For example, reading the instructions on imported packets of French coffee. And on one occasion, whilst on holiday in France, when a vineyard owner explained how he produced champagne.
When in France, the most useful phrase is 'Parlez-vous anglais?" which means 'Do you speak English?' However I once mistakenly asked a puzzled French man if he spoke French! (He did.)

Caroline Egerton, publisher.
I learnt French, German and Spanish at school and went on to study German at University. Later, I spent two months in Malay speaking countries, followed by two years in Thailand. Malay and Thai ought to feel quite different to speak from the European languages, but for some reason, when I tried to speak Malay, it was always Spanish that tried to come out. In Thailand, if I couldn't think of a particular word in Thai, it was always the German word that came into my head. I think it has something to do with the rhythms of the languages.

E. Turner, Cordon Bleu cook and caterer.
What did I learn at school? Not a lot! I studied Latin and French, but they made little impression. However, I enjoyed English, both language and literature. Reading was, and still is, a great source of pleasure.
French later became important when I took an advanced cookery course, and all the menus and specialist terms were in French. This caused dreadful problems for the Americans on the course who had only learnt Spanish at school, not French.

Appendices

Appendix 5.1.1 AS1 & AS2

Appendix 5.1 page 1 AS1

LAC Analysis sheet 1 ref: Cobuild 2 12, p9 sheet 51 processing ⇨

task	context				skills		language essentials			
1 12 languages you've learnt	2 topic	3 situation who/where	4 set	5 culture	6 receptive	7 productive	8 structure	9 lexis field/vocab/func	10 phonology	11 discourse relat/ass/123
12a"omit" 12b a) understand (2 Brits talking)	languages	L1 - self -- neutral	lone	Brit	(R)L	W	pres simple Q forms	languages -- Eng, Fch, Gk, It, get by, struggle -- understanding	-	outsider -- speaking langs -- 1
b) discuss (Brits & lang)	"	T - sts -- classroom	class	"	(R)L	S	"	" -- ", lazy -- discussng	-	equal -- speaking langs -- 1
c) understand (discuss) (Brits & lang)	"	L1 - self self - partner -- "	pair	"	LR	SW	past simple	" -- taught. learnt, study, Malay, Thai. Span, terms (= words) instruction explain -- "	-	" -- " -- 1
d) discuss (Brits & lang)	"	T - sts -- "	class	"	L(R)	S	pres simple	" -- " -- "		" -- " -- 1
e) tell (experience of learning English)	"	self - T -- neutral [homework]	lone	own	(R)	W	simple past simple pres	" -- speak, learn, willing, fairly easy, quite difficult, taught, useful -- explaining	-	sub -- " -- 1
f) discuss (sts' langs)	"	T - self self - partners -- classroom	group/ class	"	L	S	pres simple	" -- " -- finding out (survey)	-	equal -- " -- 1
review page questions & answers g) do (Q & A review)	grammar (Q forms)	author - self -- classroom	lone	Brit (pedagogy?)	R	(W)?	pres simple Q forms	" -- languages, live (vb) work (vb) -- [grammar]	-	sub -- making sentences -- 3

Appendix 5.1 page 1 AS2

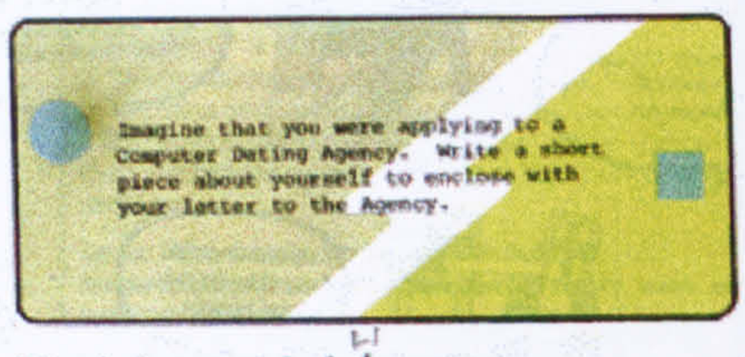
LAC Analysis sheet 2 processing

task (as col 1)	12 input	13 response	14 load	15 search	16 link	17 know	18 individual	19 demand
	/6	/5	/7	/6	/6	/6	/3	/6
a) understand (2 Brits talking)	2 simple text 9 lines, L	3 W list of names	2 L 2 x L1s 9 lines	1 L & list	1 recognise names of languages	1 names of languages	2 re speaking languages	1 W list of 8
b) discuss (Brits & languages)	2 R, T talk, sts discuss	3 S if choose to	3 class discussion	2 recall own experience	1 people & my language	1 own experience	1 own experience	4 S to class
c) understand/ (discuss) Brits & language	3 4 texts, L&R (14 lines each)	1 S to partner	4 LR 4 texts (14 lines each)	2 factual accounts given	3 indiviuals' experience	4 varied fields	2 other people's experience	2 S&W in pair
d) discuss (Brits & lang)	2 others' input	3 S if choose	2 sts' contributions	3 remember previous content	3 individuals' experience	1 just done it	2 other people's experience	4 S to class
e) tell experience of English (W)	2 phrases for W	1 homework	1 topic to W about	3 retrieve own experience	3 explain own experience	1 own experience	1 own experience	4 writing
f) discuss (sts' langs)	4 other sts' input	3 group discussion	4 group inputs	3 categorise responses	4 sts' applications	1 lang learning	1 fellow- experience	5 discuss survey accuracy
g) do (grammar Q&A)	2 answers - sts to seek Qs	2 W answers	6 guess content of Qs	5 formulate Qus	5 abstract Qus from Ans	2 common ground by now (whole unit so far)	2 not much interest	2 W sentences

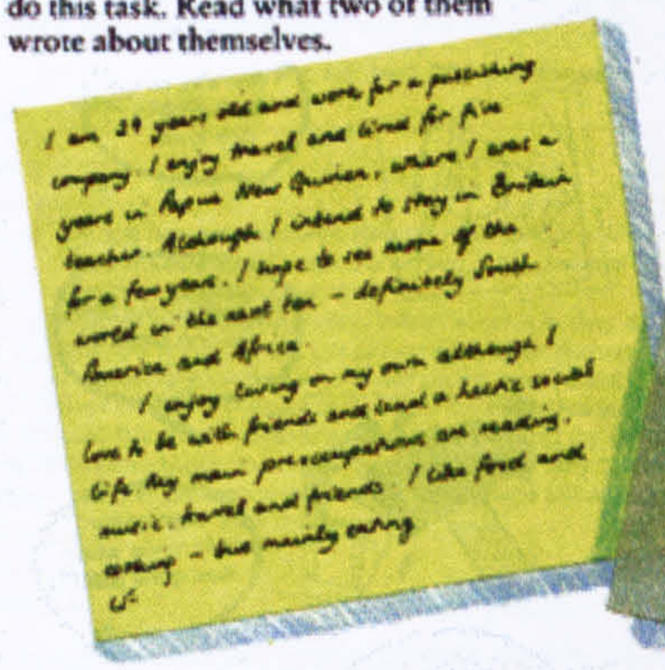
Appendix 5.1.2S

Unit 2

20 **Who would you get on with?**



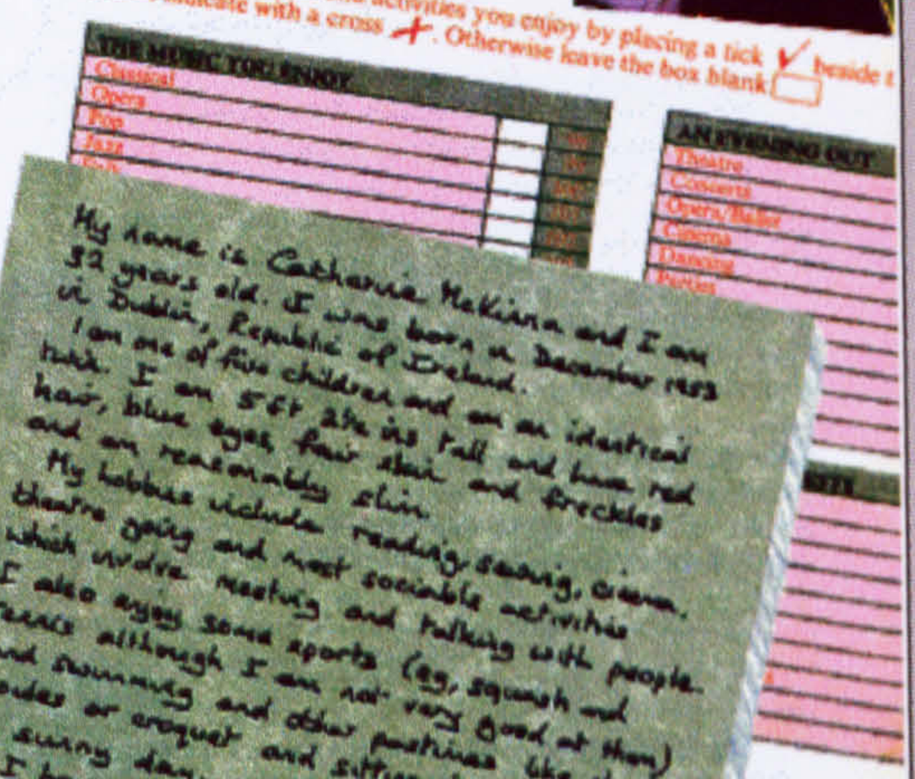
We asked everyone in the group to do this task. Read what two of them wrote about themselves.

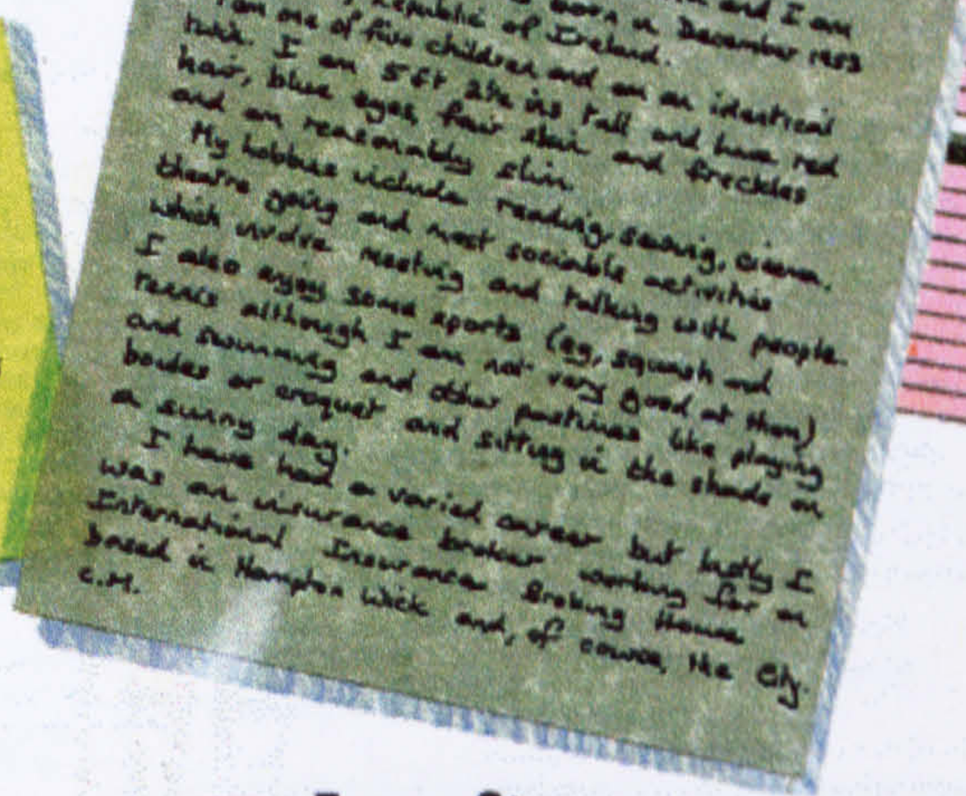

I am 29 years old and work for a publishing company. I enjoy travel and lived for five years in Papua New Guinea, where I was a teacher. Although I intend to stay in Britain for a few years, I hope to see more of the world in the next ten – definitely South America and Africa.

I enjoy living on my own although I love to be with friends and have a hectic social life. My main preoccupations are reading, music, travel and friends. I like food and eating – but mainly eating.

3. Your Interests

Indicate the interests and activities you enjoy by placing a tick ✓ beside t
please indicate with a cross ✗. Otherwise leave the box blank ☐




My name is Catherine McKinna and I am 32 years old. I was born in December 1963 in Dublin, Republic of Ireland.
I am one of five children and am an identical twin. I am 5ft 2½in tall and have red hair, blue eyes, fair skin and freckles and am reasonably slim.
My hobbies include reading, sewing, cinema, theatre going and most sociable activities which involve meeting and talking with people. I also enjoy some sports (eg. squash and tennis although I am not very good at them) and swimming and other pastimes like playing bowles or croquet and sitting in the shade on a sunny day.
I have had a varied career but lastly I was an insurance broker working for an International Insurance Broking House based in Hampton Wick and, of course, the City.
C.M.

a Think of a friend of yours who would get on well with one of these people. Explain to your partner why you think so.

20b b Listen to John and Catherine talking about what they would write.
What did Catherine say? Did she actually write what she said she would?

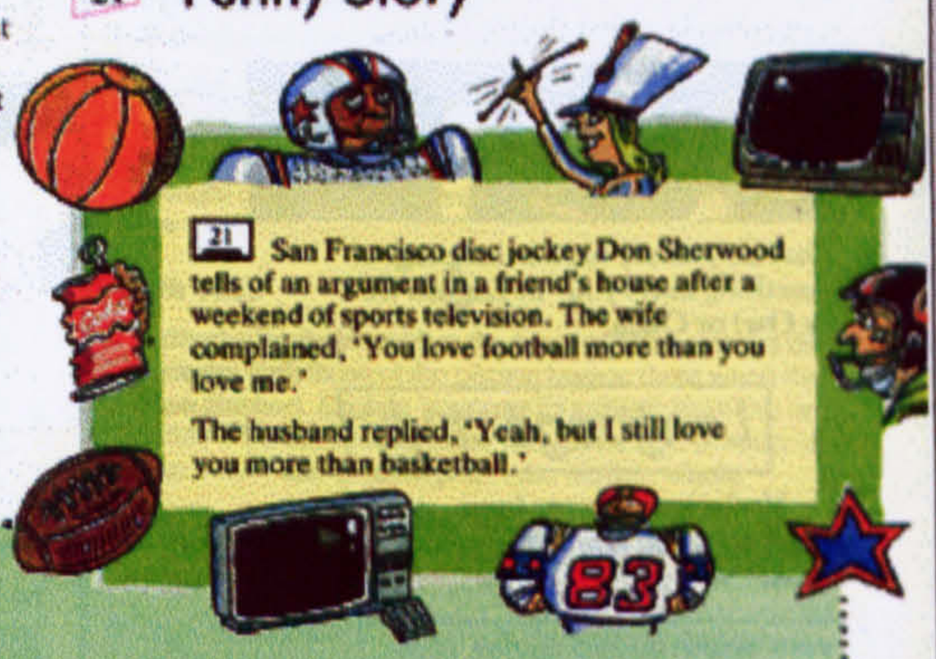
Make a list of what they have in common.

▶ Tell the class. Do you all agree? ◀

c Write about yourself. Start by talking through with your partner what you want to write.

▶ Write in a similar style to the texts above. Read each other's and comment. ◀

21 **Funny Story**



21 San Francisco disc jockey Don Sherwood tells of an argument in a friend's house after a weekend of sports television. The wife complained, 'You love football more than you love me.'

The husband replied, 'Yeah, but I still love you more than basketball.'

22 **Grammar revision**

have, has, having, had

Look at the sentences 1–10. Find:

a at least two sentences where have goes with to. What do they mean?
b two where someone has something in their character.
c two sentences describing someone.
d two where have goes with part of the verb like this: *I have lived in London for almost as long as I can remember.*
e two sentences where have means 'own' or 'possess'.

1 *We had to look in the atlas to find out where it was.* (2)
2 *I would need to meet someone who also had a sense of humour.* (20)
3 *Do you have any pets?*
4 *Do you have a favourite colour?* (128)
5 *I've never broken anything but I've had lots of stitches.* (18)
6 *I do pride myself on having a good sense of direction.* (140)
7 *She has dark hair.* (15)
8 *I have a travelcard, so I don't have to pay.* (80)
9 *I have eaten the plums that were in the icebox.* (95)
10 *Have you finished?*

In which five sentences could you use got after have?

SB

Appendix 5.1.2T

20 Who would you get on with?

Aims: 1 To section students practice in expressing likes, dislikes and the type of people they get on with.
2 To practise writing descriptions of students' own physical appearance and personalities.

Lexis: agency, argument, character [T], **characteristics, computer, funny, international, involve** [L], **love, mainly, meet/met, meeting** [L], **reasonably, shade, skin, tall**
Revision: **academic, also, cope (with), easily, intellectual, medical, precise, settle**

This section links more directly back to the first page of this unit. It begins to pull together some of the separate smaller tasks that students have done, and build on them, leading up to a full scale writing task of the same type as Catherine and the others have done.

Reading SB20a

1 The funny story in section 21 could help you set the scene, e.g. Is this a couple who have the same interests, and get on well together in all ways?

2 Help students to understand what the context of the task was – it will be more fun if they understand what a Computer Dating Agency is. Bring in an advertisement from a magazine if you can find one. Do they have them in students' countries? Remind them that they are sometimes just for friendship, not just marriage. Make it clear that the situation is a hypothetical one. Both Catherine and John (who appears later in the tape) are happily married, and are writing and discussing what they would say about themselves if they did apply to such an agency!

3 Students read Caroline's and Catherine's descriptions of themselves. Help students to ignore unimportant words and to guess meanings of words they don't know. Don't 'teach' any that are not on the list unless your students really need them.

4 Students talk informally in pairs or groups about friends of theirs who might get on with one of these people, and give their reasons. Get 1 or 2 students to tell the rest of the class what they think if they are keen to talk.
Students may produce sentences like: 'I've got a friend called . . . who smokes a lot so he/she might get along with . . .

Listening SB20b 20b*

5 Students listen to Catherine and John deciding what they would say about themselves if they were genuinely writing such a letter. If it is culturally acceptable, tell them they will be doing the same task and they will probably listen more attentively.

Key: Catherine actually wrote more than she said she would. But she missed out one thing: the fact that she hates football.

In common – from what they say on the tape:
both like talking (John says 'same here' when Catherine mentions talking to people).
Both like reading – John reads a lot.
Both hate football – John hates all sport, in fact.

6 Students look back at SB14 and see whether they guessed right about the people from the photos of them.

Key: definitely has a sense of humour JM
absolutely hates sport JM
is fairly interested in sport but hates football CM
enjoys some sports (mostly squash, tennis and swimming) CM
likes most sociable activities which involve meeting and talking with people CM
is interested in the Arts, especially cinema, and reads a lot JM

Planning SB20c PB20

7 To help students with their descriptions, ask them to include some/all of the following:
a Name, age, where born, where from
b Physical description
c Interests, hobbies
d Education, work experience
e Hopes/plans for the future

8 If students need more help, put some useful phrases like these on the board:
My name is . . . and I am . . . old. I was born in . . . in . . .
Also:
I am blonde/I have long, blonde hair.
My hobbies include . .
I also enjoy . . .
I work for/study at . .
I hope/plan to . . .

See also section 27a for some useful phrases.

Report

9 Students could write about themselves at home, then bring their reports in the following day for the others to read. Alternatively, the teacher could read a few reports to the class, and ask them to guess who the report is about.

21 Funny story

Lexis: argument, funny

SB21 21
See note 1 for TB20a for suggested procedure. The story should be read quickly and for fun.

22 Grammar revision

Aim: To give practice in the meanings and uses of **have**.

SB23 PB22

1 Students read the rubric themselves. Make sure they know what is expected of them. Explain the meaning of **have** = possess. Explain 'have something in the mind' by giving an example like 'I have a good idea' or 'He had a bad temper'.

2 Students do the activity individually or in pairs.

Key: a 1,8 (have to = must)
b 2, 4, 6
c 2, 6, 7
d 5, 9, 10
e 3, 8
Have = Have got: 2, 7, 8 3 and 4 could be rephrased 'Have you got . . .?'

Appendix 5.1.2R

20b *

JM: Er, the first thing on my list would be that erm . . . that I have a sense of humour, er, and that I would need to meet somebody who also had a sense of humour. I'm also, erm . . . I suppose, interested in the Arts. So, erm, it'd be – I'd need to – I'd want to meet somebody who was, erm, interested in the cinema and the theatre and so on. And, somebody who read a lot, because I do. And, who wasn't terribly interested in sport because I absolutely hate sport. But er –
CM: Even taking part in sport?
JM: Oh, participation in sport is even worse then actually having to watch the stuff .....
Erm. Yes. So, what er, would you put in your list?
CM: Ah, I find it very difficult to tell anybody anything about me. I suppose erm, I'd probably give them my physical characteristics, that fact that I'm, erm, not very tall. I'm five foot two and a half, and that half an inch is very important to me. And that, that I have red hair, and that I'm Irish. But probably they would get that anyway. And I would say that I'm, erm, I'm fairly interested in sport but by no means football, I couldn't meet anybody who was interested in football. Because I abhor it.
JM: Mm.
CM: Erm. And I, and I would say that I like, er meeting people. I like gregarious activities. And I like talking.
JM: Yes.
CM: And I like reading .....

21

San Francisco disc jockey Don Sherwood tells of an argument in a friend's house after a weekend of sports television. The wife complained, 'You love football more than you love me.'
The husband replied, 'Yeah, but I still love you more than basketball.'

---

23 *

CM: Well, er, I, I haven't won money. Erm, I don't think I've won – oh, I, I have won money. Er, a couple of years ago, when I was pregnant with my first child, erm, my husband put a bet on – is it the Derby? – some, some big race, erm, or the National, is that the one that's at – ?
SB: With the fences at Aintree.
CM: Yes, that's the one.
SB: The Grand National.
CM: And I won. I won on Corbière I think the horse

R

Appendices

Appendix 5.1.2 AS1 & AS2

Appendix 5.1 page 2 AS1

LAC Analysis sheet 1 ref: Cobuild 2 20, p14 sheet 52 processing ⇨

task	context				skills		language essentials			
1 Unit 2 20 Who would you get on with?	2 topic	3 situation who/where	4 set	5 culture	6 receptive	7 productive	8 structure	9 lexis field/vocab/func	10 phonology	11 discourse relat/ass/123
a) understand (preparatory rubric)	computer dating	L1 - self -- neutral	lone	Brit	R	none	pres simple pres perfect	personal -- [descrip of self] -- describing	-	equal -- personal descrips -- 1
b) discuss (who gets on with whom)	friend	L1 - self = partner -- classroom	pair - class	CHE	(R)L	S	"	" -- " -- discussing	-	" -- " -- 1
c) tell (W about self)	self	self - partner -- classroom	pair	"	L(R)	(S)W	"	" -- " -- "	-	" -- " -- 1
21 Funny story d) understand (funny story)	husband & wife	L1 - self -- neutral	class?	US	R	(laugh!)	past/pres	" -- argument, funny basketball -- understanding	-	" -- " -- 1
22 grammar review e) do (grammar: have)	grammar (have)	author - self -- classroom	lone	CHE (pedagogy)	R	tick	pres, past	" -- [everyday] -- (grammar)	-	sub -- [CHE] -- 1
23 are you a lucky person? f) discuss (winning money)	luck (money)	L1 - self self = partners - sts -- classroom	pair/ group/ class	worldwide	vL	S	pres	winning money -- win, money, lucky 2 years running -- discussing	-	equal -- glad to win -- 1
24 word power g) do (lexis exercise)	lexis (bit)	author - sts = self -- classroom	class	Brit?	vRL	S	pres	[vocab] -- bit -- [lexis]	-	sub -- sample sentences/ situations -- 1

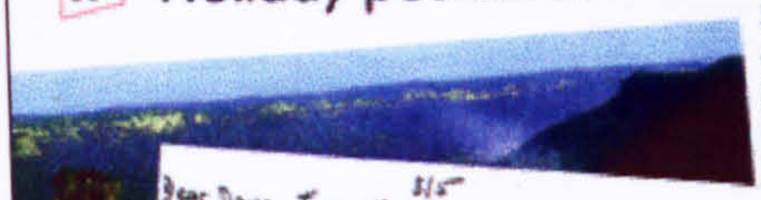
Appendix 5.1 page 2 AS2

LAC Analysis sheet 2 processing

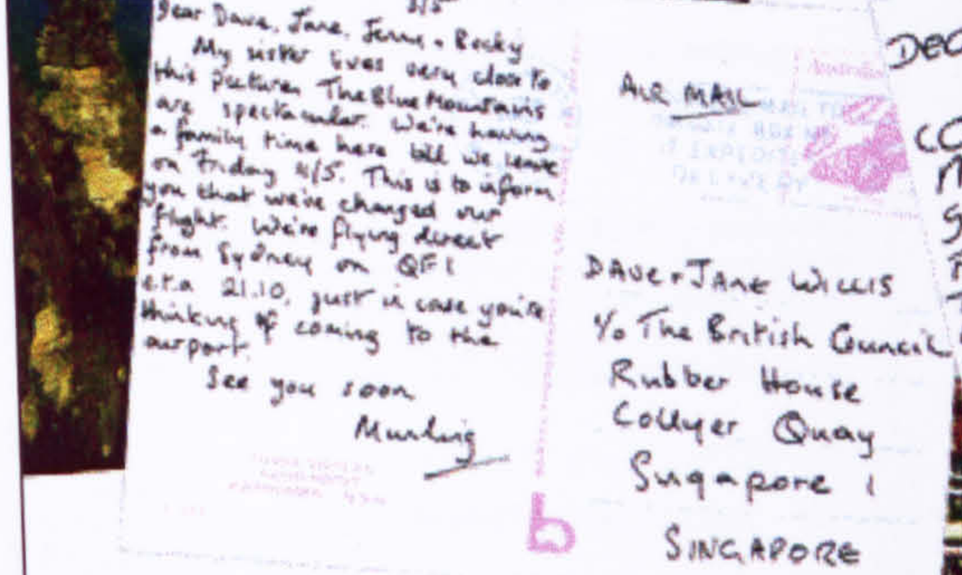
task (as col 1)	12 input	13 response	14 load	15 search	16 link	17 know	18 individual	19 demand
	/6	/5	/7	/6	/6	/6	/3	/6
a) understand (preparatory rubric)	3 R personal account	0 [none]	2 2 straight-forward accounts	2 R literally	2 common description	2 common personal description	2 just someone	0 [none]
b) discuss (who gets on with whom)	4 L L1 & partner	2 to pair & class (tog)	3 discuss pair then class	4 what would...?	3 L1's S - W S&L partner	2 descrip R,L S,L	2 just someone	3 S to partner, class
c) tell (W about self)	3 bounce own off partner	2 W description	2 own ideas, discussed	2 to hand	2 done before this unit	1 done it all in unit already	1 all about me	4 W para, some accuracy
d) understand (funny story)	2 authentic joke!	0 (laugh?)	4 2- liner	3 work it out	5 meaning of joke?	2 husband/wife relations (stereotype)	2 Ok if funny to me	2 peals of laughter?
e) do grammar: 'have'	2 simple sentences	1 own time	3 authentic sentences	3 find 'have' meanings	4 sentences: find link each time	2 CHE	3 pretty bored	1 match/ tick
f) discuss (winning money)	4 L1 chat + partner's	3 discuss with partner	2 short L1, partner	3 who done what?	2 idea of betting	2 common vice!	2 interested in winning	3 discuss
g) do (lexis exercise)	2 sentences	3 discuss in class	2 9 sentences	4 find phrases in parallel	2 words - cartoons	2 everyday?	3 bored?	4 S ideas to class

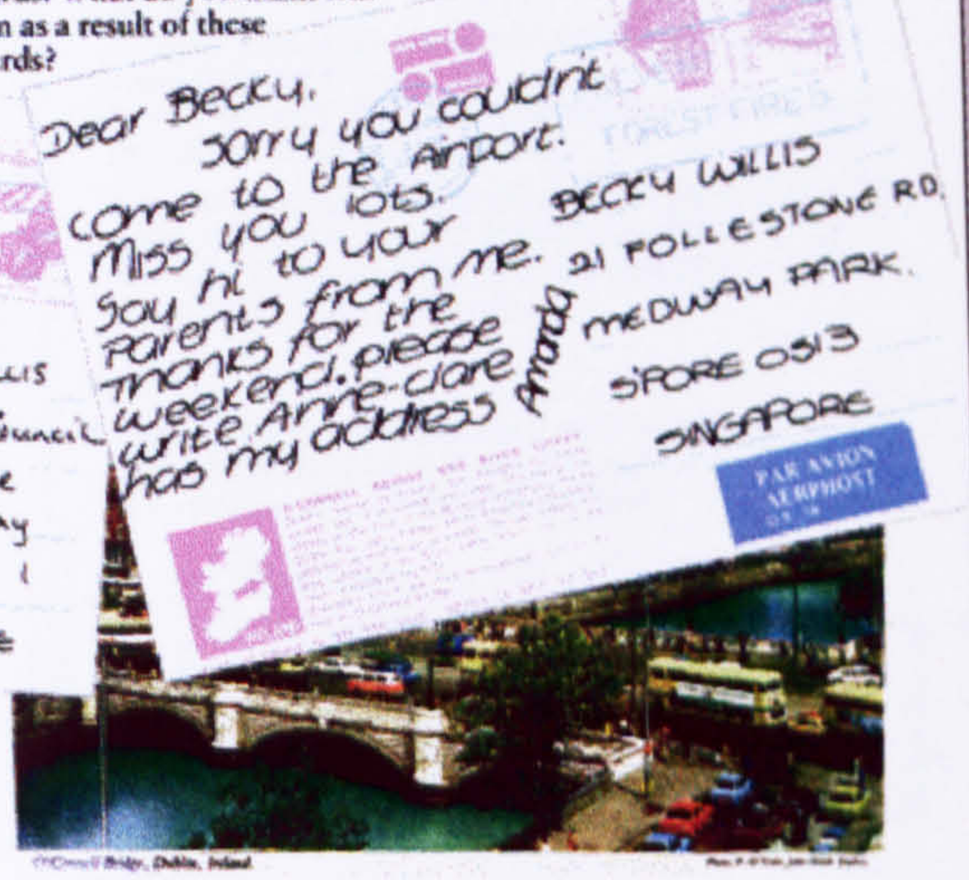
Appendix 5.1.3S

33 Holiday postcards



a Read the two postcards. Speculate about who the senders are and their reasons for sending the postcards. What do you think will happen as a result of these postcards?





b What do you think about the first postcard?

33b Listen to John and Monica. Did they say the same as you?

c What about the second postcard?

33c Listen to John and Monica again.

Why do they use these words and phrases?

Perhaps	Do you suppose
It sounds as though	I don't know
Maybe	Probably
I suppose so	

d Imagine you are on holiday abroad. You write a postcard to a friend. Your real purpose is to persuade them to meet you at the airport or station. But of course you want to do this without actually asking them directly.
Write the postcard.

34 Wordpower

time and money

1 time = minutes, hours, days, week, months etc.
Mummy and Daddy went back to Ireland some time ago. (a)
It'll rain all the time. (m)
How did they spend the time? (n)

2 to have a _____ time.
It sounds like you had a good time. (e)
We had a dreadful time.

3 time = an occasion when something happens

The second time was when we went to eat at a hamburger restaurant.
I've already boarded this flight five times and every time I ended up in Cuba.
This time our friend left a warning note. (o)

4 time = two o'clock, 8.30, 17.20 hours etc.
What time is it? Do you have the time?

5 times
The taxi cost five times as much as the bus.

Which category do these examples belong to?

(a) Took a very long time getting there. (n)
(b) People we met on it said they'd been many times before. (n)
(c) Look at the time. We're going to be late.
(d) But next time I knew I could do it. (m)
(e) We had a very busy time at work last week.
(f) My job is a hundred times more difficult than playing the piano. (m)

Spend goes with money as well as time. What other words are like spend?

(g) That's very expensive. We can't afford it.
(h) You'd waste a lot of time and you'd feel exhausted when you got there. (n)
(i) Business is bad. We're losing quite a bit.
(j) How much did you actually spend?
(k) If you go by taxi you'll save ten minutes or so.
(l) We are saving for our holidays.

SB

Appendix 5.1.3T

33 Holiday postcards

Aims: 1 To practise making and understanding an implied request.
2 To practise reading and writing holiday correspondence.
3 To focus on the meaning and use of 'hedging' words and phrases, e.g. **perhaps, I suppose**.

Lexis: country [T], **miss (you lots)**, **purpose**, **reason**
Revision: **basically**

SB33a PB33

1 After students have read the postcards and speculated on their meanings in pairs, put some of their suggestions on the board. Don't confirm or give away any answers at this stage – this will be the point of further activities. It may help the students to speculate if you put a framework on the board:

Place:
Relationship Munling – Willises/Amanda – Becky:
Holiday?:

Result:

SB33b 33b*

2 Students listen to John and Monica to compare and/or confirm their answers.

Key: Postcard 1
John and Monica guessed correctly that Munling is a friend of the Willises on holiday in Australia, and that she sent the postcard in the hope that they would meet her and her family at the airport.
There is an additional fact that John and Monica did not know. Munling had left her car with the Willises while on holiday and they had offered to meet the family at the airport. The postcard is, therefore, by way of a reminder.

Postcard 2
They were correct in that Amanda was a friend of Becky's; that Becky wanted but was unable to see her off at the airport; that Amanda had gone back to live in Ireland. They did not know that Amanda had spent her final weekend in Singapore with Becky's family.

Listening SB33c 33c*

3 In pairs students speculate about why John and Monica have used the words and phrases in the box.

Key: John and Monica say 'perhaps', 'it sounds as though' etc. because they are not sure; that is, they are speculating or guessing.

Planning SB33d

4 Elicit a few ideas about how to drop hints, e.g.
We don't get back till 10.30 at night. I'm afraid the children will be very tired.
I hope we don't have to wait too long for a taxi/bus at the airport.
It will be nice to get home but I always hate the long busride from the airport.
I wish we'd left our car in the airport carpark, but it's so expensive we couldn't afford to.

It is probably better not to put these on the board in full, otherwise all you have left is a mechanical copying exercise. In groups students write their postcards. Go round and help as necessary.

Report

5 Take finished postcards in and read them out to the class. Ask students to decide how persuasive and subtle they are – a scale of 1 to 5 might be fun here.

34 Wordpower

Aim: To practise recognising words which can be used both with expressions concerning money and expressions concerning time.

Lexis: afford, lose, loss [T], **spend, time, waste**

SB34 PB34

1 In pairs students match sentences a–f to the correct lexical categories. Refer them to the Lexicon for further explanation of the categories.

2 Students find the words in sentences g–l which can be used with both time and money.

Key: a 1 b 3 c 4 d 3 e 2 f 5
g afford
h waste
i lose
j spend
k save
l save

3 Ask students if there are words in their own language which can be used with both time and money. In English there is a saying, 'Time is money.' Do they have a similar expression?

Appendix 5.1.3R

13b

JM: Then we move, move on to Postcard Speculations. I'll read this one out. Er. There's a postcard for us to look at.
MJ: Are you supposed to read the back as well?
JM: Mm. Read the two postcards. Note where they come from. Spec- speculate about who the senders are, and their reasons for sending the postcards. Do you think they are both away on holiday? What do you think will happen as a result of each of these postcards?
MJ: Can I see that?
JM: 'Dear Dave, Jane, Jenny and Becky,' mine says. 'My sister lives very close to this picture. The Blue Mountains are spectacular. We're having a family time here till we leave on Friday the eleventh. This is to inform you that we've changed our flight. We're flying out on – from Sydney on QF1 E.T.A. 21.10 just in case you, you're thinking of coming to the airport. See you soon,' and then I can't read the signature. So, this one is obviously –
MJ: It's a friend of theirs that's gone to Australia? The Blue Mountains are in Australia, aren't they?
JM: That's right. Yes.
MJ: And erm, erm . . . The purpose of the postcard is, er . . . to tell, erm, the Willises about the holiday and basically to suggest that they might like to come and pick them up.
JM: Yes. It suggests – 'If you are thinking of' rather than, er, 'please be at the airport'. But yes, I imagine, erm, they probably will turn up. So what –
MJ: I can't read the name
JM: No, I can't read the name either.

13c

MJ: Erm, the other postcard is a picture of, er, a bridge in Dublin, and, erm, it's written to Becky Willis and it says, 'Sorry you couldn't come to the airport. Miss you lots. Say hi to your parents from me. Thanks for the weekend. Please write. Anne Claire has my address. Amanda.'
JM: Mm.
MJ: Perhaps this, erm, this is a friend of Becky's who, er, left, erm, and was hoping Becky would be able to come to the airport and say goodbye to her.
JM: Mm. So it sounds as though, either maybe she lives in Ireland and, and er, has gone back, or has just moved to Ireland and er, won't be back for some time.
MJ: Mm.
JM: With the invitation to write.
MJ: It's, well, yes, I suppose so. 'Thanks for the weekend.' Do you suppose this, erm, this girl went to Singapore for the weekend from Ireland?
JM: Singapore?
MJ: Yes, it's written to Becky in Singapore.
JM: Oh, I see. Oh, erm. Mm.
MJ: I don't know.
JM: No, perhaps they, er –
MJ: I think we move on from here.
MJ: Erm, as a result of this postcard, erm, I think Becky will write back.
JM: Yes, probably. If she's still in Singapore or wherever.
MJ: Right. I think that's the end of that one.
JM: Okay.

R

Appendices

Appendix 5.1.3 AS1 & AS2

Appendix 5.1 page 3 AS1

LAC Analysis sheet 1 ref: Cobuild 2 33, p19 sheet 53 processing ⇨

task	context				skills		language essentials			
1	2 topic	3 situation who/where	4 set	5 culture	6 receptive	7 productive	8 structure	9 lexis field/vocab/func	10 phonology	11 discourse relat/ass/123
33 holiday postcards										
a) understand (postcards & reasons for)	postcards	originator - L1 - self - partner -- classroom	pair, class	WEuropean	RvisL	S	pres, neg, past, modal (cld), fut, pres perf	holidays -- spectacular, flight, airport, weekend -- telling news	-	outsider -- postcard writing -- 2
b) invent (a postcard text)	"	self - author -- neutral	lone	own	(R)	W	"	" -- " + own -- hinting	-	sub -- " -- 2
34 [word power] time & money										
c) do (vocab exercise)	time & money	author - self = partner -- classroom	pair	Brit	R	tick	past	(uses of) time vocab -- misc (sentences) -- [learning vocab]	-	sub/equal -- learning -- 1
35 understanding the Pennines										
d) understand (brochure language)	holiday	originator - self = partner -- classroom	lone/ pair	"	Rvis	S	pres	holidays -- paradise,magnificent fishing, angler, moor pheasant, hare, accommod'n,meal... -- suasion	-	equal -- holiday brochure style -- 3
e) complete (phrases to make texts)	"	author - self = partner - sts -- classroom	lone - pair - class	own	R, L	W, S	pres, demonstraive	" -- no better, excellent, magnificent, enjoyed, facilities -- describing	-	sub -- " -- 3
f) do (phrase building)	lexis	author - self - partner -- classroom	lone - pair	pedagogy	RL	S	we did... that's how... past	misc vocab -- [holiday activities] did -- learning	(inton'n & stress of phrases?)	sub -- learning -- 1
g) do (grammar revision)	grammar (been)	self - partner pair - sts -- classroom	pair - class	"	R	S	been	grammar -- been, been -- learning	-	sub -- learning -- 1

Appendix 5.1 page 3 AS2

LAC Analysis sheet 2 processing

task (as col 1)	12 input	13 response	14 load	15 search	16 link	17 know	18 individual	19 demand
	/6	/5	/7	/6	/6	/6	/3	/6
a) understand (postcards & reasons)	4 R 2 postcards	3 discuss implications, pair, class	2 2 postcards + discussion	2 implications of 2 postcard contents	4 postcards' implications	2 travel, holidays	1 intriguing: why?	2 discuss partner
b) invent (postcard text)	1 rubric alone	2 W: follow text for model	1 1 input (rubric)	5 find story, persuade	4 express ideas in W	1 own story	1 own story	4 W postcard
c) do (vocab exercise)	2 examples of 'time'	2 study, tick	4 11 sentences, match with 4 concepts	2 apply rule	2 match sentences to concepts	2 CHE	3 just boring, bits	1 tick
d) understand (brochure language)	5 travel brochure (tough text)	2 chat with partner	5 30 lines of text, 7 pictures, guess words	5 some guesswork	4 conceptualise brochure style	4 connect with 'real life'	1 places, people, pictures	3 discuss & report
e) complete (phrases to make text)	2 phrases	2 discuss with partner	2 13 phrases	4 find facts to fit	2 attach own to phrases	2 own choice	1 own choice	4 discuss/ tell class
f) do (phrase building)	1 phrases to learn	3 repeat & learn	3 13 phrases (seriatim)	2 repeat with minimum thought	3 relate examples to uses	2 common task	3 dead bore	2 repeat
g) do (grammar revision)	2 sentences, work on	3 report to T, to class	2 10 sentences	2 find pattern	3 compare 2 uses of 'been'	2 misc 'CHE'	2 work out pattern	4 discuss & report

Appendices

Appendix 5.1.4S

Unit 4

43 Best paid and worst paid

a Look at the list of jobs below.
In groups, discuss which are the three best paid jobs and the two worst paid in your country. Write a list.
Do you think that any of the jobs on the list are overpaid or underpaid in your country?
Have you any idea which are the best and worst paid in Britain?

a nurse in a hospital
a miner at the coalface
a shop assistant
a car assembly worker in a factory
a bank manager
a dentist
a schoolteacher
a plumber
a top professional footballer
a nuclear scientist

▶ Tell the class what you have decided. ◀

Listen to the other groups and see if they think the same as you.
How many groups had the same person at the top of their lists?

43b **b** What did Caroline and Stephen think? Draw a table like this one, and fill it in after you have listened.

Best paid	.....
Worst paid	.....
Overpaid	.....
Underpaid	.....

Compare their opinions with:

what you think about your country.
what you decided about Britain.

44 Language study

44a **a** Listen to Caroline and Stephen talking about the schoolteacher and the nurse. How many times do they use the words **say, think** and **know**? What phrases do they use with them?

b How many ways could you categorise these phrases?

paid reasonably well
get a lot of money
Too much!
paid a lot
the lowest
are underpaid
were the worst paid
gets even less
get well paid
the same pay
get less pay
earn more than just £1.20 an hour
at least £108 a week in wages
the salary is just not enough
can earn as much as $200 a day
a low income
a highly paid job
gets a good salary

45 Two jokes

a At the dentist's

Do you like going to the dentist? Why not? How much does it cost (roughly) to have a tooth out in your country?

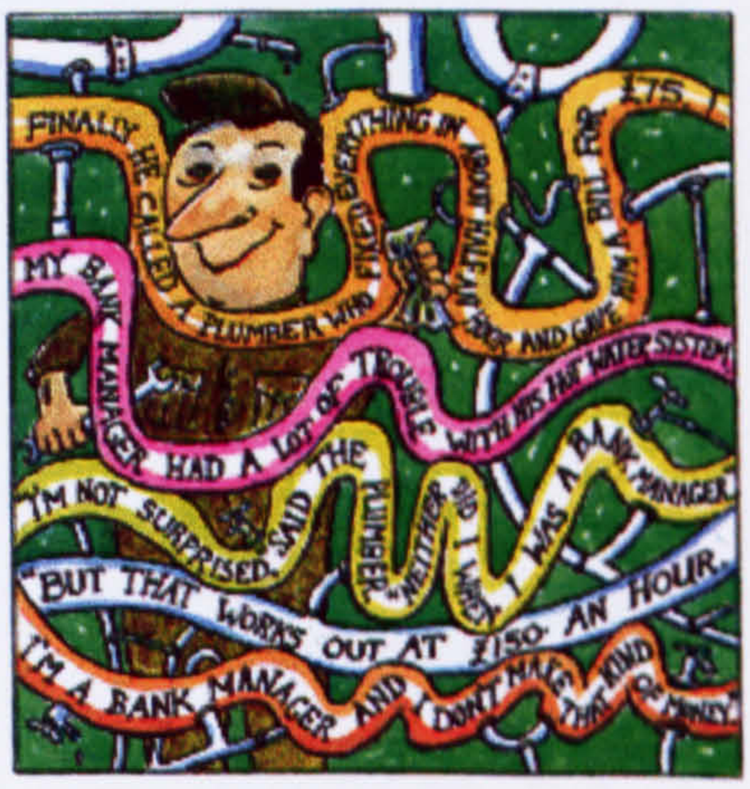
Compare prices with other students.

What could be the last four words of this joke? Decide what you think they are then tell the class.

> A woman went to a dentist in Baghdad to have a tooth out and was told it would cost the equivalent of £30.
> 'But that's ridiculous!' she said. 'My husband has to work two hours for that.'
> 'Madam,' the dentist replied, 'If you like, I _____ _____ _____ _____.'

b If you call the plumber ...

The five parts of this joke have been mixed up. Can you work out the joke?



45c **c** Listen to the jokes.

46 When I was a paperboy

What kind of work can schoolchildren or young students do in your country to make some money? Tell each other.

Caroline and Stephen talk about their jobs as baby-sitter, paperboy and factory-worker. Stephen also mentions a friend who's a waitress.

Put their jobs in order from worst paid to best paid.

46 Listen and see if you were right.

SB

Appendices

Appendix 5.1.4T

SB43 PB43

Aims: 1 To provide practice in the skills of discussion, comparison and evaluation.
2 To provide practice in the skills involved in reaching a cooperative decision.

Lexis: anybody, handle low(est), nuclear, nurse, professional [L], **science** [T], **worst**
Revision: **definitely, high, less, less than, money, more than, must be/must get, no idea, over/underpaid, probably, top, worth**

Task SB43a

1 In groups students discuss the questions and note their answers. Encourage them to use a similar table to 43b.

2 Check that students understand the difference between best paid and overpaid (paid too much for what they do).

If the class is mixed nationality, you could put them in groups according to where they come from. Also, if some jobs don't exist in some students' countries, get them to substitute similar ones e.g. quarry worker for coal miner.

3 Students report back briefly. Encourage them to give reasons for their opinions, e.g. Coal miners should be better paid because it is dangerous and unpleasant work.

Listening SB43b 43b*

4 Students listen to the tape of Stephen and Caroline and complete the table in groups.

Key: Best paid: footballer, bank manager, nuclear scientist, (possibly dentist)
Worst paid: shop assistant, car assembly worker
Overpaid: footballer, bank manager
Underpaid: teachers, nurses
Stephen and Caroline are more or less right, apart from the footballer; they are like actors – the best of them are very well paid, the worst very badly.

5 Get groups to tell you informally how their opinions compare with what they heard on the tape.

SB44 PB43

Aims: 1 To study the common uses in discourse of phrases with the verbs: **say, think** and **know**.
2 Understanding phrases expressing amount or comparison in connection with money or pay.

Lexis: high(ly), income, low, nurse, partner, wages, worst

SB44a 44a* PB43

1 Students refer to their transcripts to find the words and phrases **say, think** and **know** occur in.

Key: Caroline uses **think**: Dentists are well paid, I **think**.
Caroline uses **know**: It's hard to **know**, isn't it?
Stephen uses **say**: Who are both I would **say** underpaid. I'd **say** the shop assistant and the car-assembly worker were probably the two worst paid.
Stephen uses **think**: But I **think** a shop assistant probably gets even less. I **think** the footballers are probably overpaid. I **think** bank managers are overpaid . . .
Stephen uses 'know' in: I don't **know** whether it's . . .

SB44b

2 In pairs students find as many ways of categorising the phrases as possible. (They all express a degree of quantity or amount of money or pay.)
Possible ways: **a** paid a lot / not paid much / reasonable / other (e.g. the same) **b** paid too much / not paid enough / reasonable/neutral
Students could add to this list and make pairs of phrases with opposite meanings: e.g. too much, not enough / at least, at the most etc.

SB45 PB45

Aims: 1 To give practice in understanding narrative.
2 To give students practice in listening for the order of events in a narrative and for specific information within the narrative.

Lexis: dentist, equivalent (of), fixed joke, neither (did I), surprised, system, trouble [L]
Revision: **ridiculous**

SB45a 45a

1 Set the scene: lead the discussion about going to the dentist's.
Discussion points: ask students if they have much trouble with their teeth. How often do they go to the dentist? How much does it cost? Does the Government help pay for treatment? (In Britain, children under 16 or in full-time education get free treatment.) Do they enjoy going to the dentist? If not why not?

2 Students read the story and try to work out the last 4 words for themselves, then tell each other their answers.

3 Play the tape which has the whole story with the original ending.

SB45b 45b

4 Make sure students know what a plumber does. Ask if it is highly skilled work; do you need to be highly educated to be a plumber? How much do plumbers charge in the students' countries for a house call?

5 Students do the puzzle in pairs.

6 Students listen to the tape to see if they have the same order.
Discussion point: you could ask the students for stories about plumbers or any similar experiences they have had when they think they have been overcharged. Have they ever had trouble with any water systems where they live?

SB46 PB46

Aims: 1 To practise the skills of discussion, comparison and evaluation (of children's and younger students' jobs).
2 Structure: if it wasn't for the tips, she wouldn't be getting much at all. If it wasn't for the . . . = Without the . . . If she didn't get the . . .

Lexis: dollars, factory, low, tend, waiter, waitress
Revision: **mainly, pretty awful**

Task SB46

1 In groups students tell one another about their experiences and list the types of work younger people can do to make money.

2 Students speculate about how to order the 4 jobs in terms of pay. (It's a cake factory, if they want to know.)

Listening SB46 46*

3 Before they listen to the tape, warn students it won't be absolutely clear from what Stephen and Caroline say exactly which order the jobs come in.

4 Suggest students make a note of how they get paid for each of the 4 jobs. Also ask them which person they think did which job.

Key: Stephen was a paperboy; he didn't say how much he was paid, but he said it was the lowest paid work he did.
Stephen also did babysitting for 6 pence an hour when he was eight! This was paid even less.
Caroline worked in a cake factory for not much more than a pound an hour.
Stephen's friend works as a waitress and claims she can sometimes get as much as $200 in a day (including tips – i.e. her basic pay is probably very little).

Appendix 5.1.4R

UNIT 4

41b

SB: Look at this list of jobs. Decide which three of these jobs are the best paid in Britain, and which two are the worst paid. Do you think any of them are underpaid? So we have a nurse in a hospital, a miner at the coalface, a shop assistant, a car-assembly worker, a bank manager, a dentist, a schoolteacher, a plumber, a top professional footballer, and a nuclear scientist.
CF: Mm.
SB: Well. I reckon a nuclear scientist is paid reasonably well.
CF: What about a professional footballer? They get a lot of money don't they?
SB: Too much.
CF: Mm.
SB: Yes, that must be top of the list. Bank manager must also be pretty high up.
CF: Got no idea.
SB: Anybody who handles money tends to get paid a lot.
CF: Mm. What about the lowest? Schoolteacher.
SB: A schoolteacher, or a nurse.
CF: Mm.
SB: Who are both I would say underpaid. But I think a shop assistant probably gets even less. I don't know whether it's – I'd say the shop assistant and the car-assembly worker were probably the two worst paid. But what about the dentist? He must get well paid.
CF: Mm. Dentists are well paid, I think. It's hard to know isn't it?
SB: Yeah. They're probably worth the money though. I think – I think the footballers are probably overpaid and I think – I think bank managers are overpaid, but then, anybody who's paid more than me is probably overpaid.
CF: So we both agree that teachers and nurses are underpaid.
SB: Teachers and nurses are definitely underpaid.
CF: Mm.

14a

CF: Mm. What about the lowest? Schoolteacher.
SB: A schoolteacher, or a nurse.
CF: Mm.
SB: Who are both I would say underpaid. But I think a shop assistant probably gets even less. I don't know whether it's – I'd say the shop assistant and the car-assembly worker were probably the two worst paid. But what about the dentist? He must get well paid.
CF: Mm. Dentists are well paid, I think. It's hard to know isn't it?
SB: Yeah. They're probably worth the money though. I think – I think the footballers are probably overpaid and I think – I think bank managers are overpaid, but then, anybody who's paid more than me is probably overpaid.
CF: So we both agree that teachers and nurses are underpaid.
SB: Teachers and nurses are definitely underpaid.
CF: Mm.

45c

A woman went to a dentist in Baghdad to have a tooth out and was told it would be the equivalent of £30. 'But that's ridiculous!' she said. 'My husband has to work two hours for that.'
'Madam,' the dentist replied, 'If you like, I will take two hours.'

Liked that one did you, yeah, well here's another, here's another.

My bank manager had a lot of trouble with his hot water system. Finally he called a plumber who fixed everything in about half an hour and gave him a bill for £75. 'But that works out at £150 an hour. I'm a bank manager and I don't make that kind of money.'
'I'm not surprised,' said the plumber. 'Neither did I when I was a bank manager.'

46

CF: What's the lowest-paid job you've ever had?
SB: Probably way back when I was a paperboy. Or babysitting maybe. That's sixpence an hour –
CF: How old were you then?
SB: Must have been about eight.
CF: Babysitting at eight?
SB: Well, little sisters were even smaller than me, so –
CF: Yes, I think my worst pay was in a factory – cake factory when I was a student and I can't remember how much it was, it wasn't much more than a pound an hour. But it was pretty awful.
SB: Mm. I, I've got a friend who's a waitress who claims that she gets about twenty dollars an hour including tips.
CF: Mm.
SB: As a waitress. Which is, what, twelve pounds an hour.
CF: Mm.
SB: And sometimes even as much as two hundred dollars in a day.
CF: So that's mainly from tips.
SB: That's mainly from tips. I think if it wasn't for the tips she wouldn't be getting much money at all.
CF: Yeah.
SB: And I think that might be where women have a slight advantage. I think they tend to get tipped more generously than men.

R

Appendices

Appendix 5.1.4 AS1 & AS2

Appendix 5.1 page 4 AS1

LAC Analysis sheet 1 ref: Cobuild 2 43, p24 sheet 54 processing ⇨

task	context				skills		language essentials				
1 43 best paid & worst paid	2 topic	3 situation who/where	4 set	5 culture	6 receptive	7 productive	8 structure	9 lexis field/vocab/func	10 phonology	11 discourse relat/ass/123	
a) discuss (jobs & pay)	jobs & pay	self - partner - sts -- classroom	group, class	own, Brit	(R)L	(W)S	comparatives, superlatives, pres	job titles -- [professions] pay, wages, salary -- discussing	-	equal -- sociology of jobs -- 2	
44 language study b) find (words in text)	"	L1 - self - sts -- classroom	class	worldwide	R [L text]	S	"	opinions -- say, think, know -- understanding	-	equal -- speaking langs -- 1	
c) discuss (categorisation of phrases)	comparing pay levels	author - self - partner -- "	pair	"	RL	S	"	pay -- paid, well, worst paid, get (£) -- discussing	-	sub, equal -- describing levels of pay -- 2	
45 two jokes d) understand (jokes)	dentist	originator - self -- neutral	lone	CHE	R	S	past	dentists -- dentist, have (tooth) out, equivalent, work -- guessing words		equal -- jokes -- 3	
e) understand (a joke) + discuss	plumber	" -- "	pair	worldwide	visR, L	S	"	plumber -- bill, system, bank manager, work out at -- reordering text	-	" -- " -- "	
46 When I was a paper boy f) discuss (children's jobs)	children's jobs	self - partners -- classroom	group	"	R(L)	S	pres numbers	pay for job -- awful, paperboy, lad, babysitting, factory, waitress -- comparing	-	equal -- jobs, pay -- 1	
47 equal opportunities g) understand, discuss	woman working in bakery	originator - self, partner -- classroom	lone, pair	Brit	R⊥	S	pres, pres part, numbers	job in baker's -- bakery, break, oven, weight, take home -- (complaining) discussing		equal x 2 -- socio: wages & hours -- 2	

Appendix 5.1 page 4 AS2

LAC Analysis sheet 2 processing

task (as col 1)	12 input	13 response	14 load	15 search	16 link	17 know	18 individual	19 demand
	/6	/5	/7	/6	/6	/6	/3	/6
a) discuss (jobs & pay)	2 other's ideas on jobs	3 discuss in groups	2 group discussion	2 think about jobs	2 common words for common jobs	2 jobs	2 who does what, own country incl	3 discuss & report back
b) find (words in text)	2 R transcript	1 find words in transcript	2 find words in short transcript (16 lines)	2 find words in transcript	4 say/tell	1 heard it, now R it	3 so what?	1 find words in transcript
c) discuss (categorisation of phrases)	2 R phrases	2 discuss with partner	2 R 18 phrases	5 categorise (higher level organisation)	4 differences of meaning	2 description of pay differences	2 experience similar?	3 find categories
d) understand (joke), discuss	2 joke	2 discuss with partner	1 R 6 lines	5 find key words for joke	4 joke	1 dentists	2 identification with dental patient	5 key words of joke
e) understand (joke), discuss	3 R jumbled text	2 discuss with partner	2 R 5 lines embedded in visual	3 find right order for phrases	4 joke	2 plumbers	3 plumber & bank manager	4 order joke text
f) discuss (children's jobs)	3 L L1 narrative, partner	3 discuss with partner	3 L 21-line dialogue, partner	4 L & judge likelihood	3 children's jobs	2 jobs	2 experience similar	3 discussion with partner
g) understand, discuss (jobs, pay)	4 R complaint re £/hours	2 discuss with partner	2 R 16 lines L partner	4 figures	3 understand job situation	3 jobs & pay	2 socio problem	3 discussion with partner

Appendices

Appendix 6.1

Experimental materials

This appendix consists of three examples of experimental materials. The first, *A visitor*, was an illustration for discussion at a workshop with teachers in Perú. The second and third were written for the researchers own students as simulations: *Travel award* and *Alton Towers*.

Appendices

Appendix 6.1.1

Interactive exercise from Perú

A visitor

1. Situation

An English-speaking person, Mr/Mrs Robinson, is visiting Lima on business and is bringing his/her son John (aged 16) with him/her. He/she writes to Mr Brown* and asks if he can suggest what John should do while he is here. Mr Brown decides to ask the students for ideas.

2. Sequence

a. Teacher reads Mr/Mrs Robinson's letter to the class. It includes information about John's interests (eg surfing, dancing, meeting other students, Peruvian food, buying something to take back for his girl friend... - whatever teenagers are interested in!). Students take notes, individually.
b. Students read, individually, a variety of material in English about things related to John's interests and beyond (what else should he do/see while he is here?).
c. Students discuss in groups of four what they think would be best for John and write a programme for him for the weekend, Saturday & Sunday, making notes and agreeing.
d. Students each individually write a letter to John explaining what they suggest. Mr Brown has agreed to send him the best one!

3. Materials

a. The letter from Mr/Mrs Robinson.
b. At least four short texts about possibilities.
c. A blank diary to help students fill in the activities.

4. Levels

The letter from Mr/Mrs Robinson can be more complicated, or less, eg mixing up John's interests with other things about him which are irrelevant; or it could be just a straight list. The reading can be long and difficult or short and easy, and the timetable could cover three days instead of two, or only one.

(*Note: Mr Brown was the actual Director of the Centre.)

Appendix 6.1.2

Travel award

The Smith Travel Award: a simulation

The Smith Foundation is offering a travel award which will allow two people to spend four weeks anywhere in the world for any purpose they are interested in. All projects must start within six months of the award. Past winners of the award have:
- gone painting in the Kalahari desert
- driven a relief lorry in Bosnia
- learnt how beach lifeguards are organised in Australia.

There are two sets of participants: applicants and interviewers. The tasks are:
1 *For applicants:*
Work with a partner. Your are applying together for the travel award. You must think of a suitable project and write a short statement about it.
- Say where you would like to go and why;
- Explain what you would do and how you would do it;
- State what you hope to achieve;
- List the preparations you will make over the next six months;
- Say why you think you are suitable people to do this project.

Write no more than half a page.
2 *For interviewers:*
Work in a group. You are going to interview the applicants for the award. You will have four pairs of applicants to interview and must make the award to the best project. Discuss and decide on what criteria you will use to help you decide, for example:
- the characters of the applicants;
- the value of the project;
- the likely result;
- value for money.

Examples of possible projects include:
- designing a solar water pump for Somalia
- finding out how South American Indians use plants for healing
- making a film in Armenia
- studying bees in Greece
- learning about whales in Antarctica
- working in a conservation area in Europe
- finding out if the market economy has anything to learn from an Arab souk
- photographing the next Olympics
- ...

But try to find your own.

Useful phrases for applicants:	Useful phrases for interviewers:
We should like to...	why do you want to...?
It would be useful/interesting because...	what is the point of...?
It is necessary/important/vital for... to...	what would you do if...?
If we were given the award we would...	what would you do with...?
The result would be...	

Timing:
Introduction: 15 minutes; preparation: 20 minutes; each interview: 10 minutes, including reading the application, the interview, making the decision (grading?); Results and discussion: 15 minutes.

Appendix 6.1.3

Alton Towers

Visit to Alton Towers: a simulation

1. Situation

You have been asked by [N - the school trips organiser] to produce a programme for a visit to Alton Towers for students from the school. The article 'The agony and the ecstasy' [review of Alton Towers from recent edition of a national newspaper] will tell you all you need to know about this theme park, but you will also need other information, like how far from here it is.

2. Demands

You are asked to produce:

1. An administrative plan for the day, giving times of departure, where to meet, how many can go, what transport you will use, who will drive... and so on. Think about all the things which will have to be planned in advance, and try to make arrangements for them.
2. A poster for the classroom notice boards giving all the necessary information to students (what it is, what is exciting about it, when, where, cost...). The poster for [the local entertainments centre] and the picture with the newspaper article may give you some ideas.
3. A suggested timetable for the visit. You should take into account the recommendations made by [the author of the article], but remember that there may be two parties going from [the school] and they cannot both get there at the same time. What should be visited first? What later? When and what will students eat/drink, and where?
4. A letter which is to be faxed to the parents of the younger students asking permission for them to go. The letter should explain how useful the visit will be for them *educationally*.

3. Resources

Most of the information you will need is in the newspaper article, but there may also be other things you will want to know. Also, you may need practical resources like coloured pens and paper for the poster. I may be able to help: ask me if you want anything.

4. Organisation

Work in three groups, with four students in each group. Elect a **leader** for your group. Then agree who will be **secretary**, to write notes as you discuss what you want to do; **messenger**, to find out more information when you need it; and **artist**, to design the poster. All decisions should be agreed by the group as a whole, but if you cannot agree, the leader has to decide what will be done.

5. Tasks

a. The first task is to plan your time. You have until 1030 to do all the other tasks below. Don't forget to include time for checking the work when you have finished it and time for reporting back to the other groups. Discuss what there is to do, and work out a timetable for the rest of the lesson.

b. Then all four of you read the article and discuss the information it contains. Make sure you understand it all (use dictionaries if necessary). Then cross out any parts of it which seem irrelevant for the trip.

c. Decide how many jobs there are to do, and share out the work between you. For example one pair could write a draft plan, another pair could make recommendations about food, ... - or you could work individually, if you prefer.

d. Discuss all the parts in draft to make sure you agree them as a group.

e. Present your ideas to the other groups - remember to leave time for them to present their ideas to you!

f. The final task is to decide which is the best of the three programmes.

Appendices

Appendix 6.2

Scenarios for trial series

The seven scenarios produced for the trial series illustrate a range of exploitations of a single format. Some are accompanied by realia - advertisements, postcards, reproductions of faces or sculptures, books; some are inventions based on actual texts; and one is pure instruction, with the students providing all the information. The following notes give some background information to the scenario materials.

1. Introduction to Dorset

The students had already been at the school for some weeks, and so were familiar with the area and events in it which might be of interest to a those in a new intake. They were each given two of the colour pictures of faces from which to choose, and the whole set of advertisements, which were chosen from a local newspaper to stimulate interest.

2. A job in Dorset

The advertisements were again taken from a local newspaper and some were deliberately rather odd (mushroom harvester, animal assistant) so as to give scope for creative responses in Task 2. The realism of this scenario was somewhat compromised by the fact that most students would probably be perfectly able to afford to stay on without extra funds if they wanted to.

3. Compatible friends

A revised version of Holiday acquaintances. The note sheet now includes the rubrics and is intended to be self-explanatory so as to reduce the animateur role for the teacher. There were 4 picture postcards with an invented message on the back, written in various forms of handwriting, each reproduced 3 times - ie 12 postcards in all - so that there was one for each student. The destabilising challenge cards were only rarely used but caused some amusement when they were.

4. Lecturer

A scenario requiring labour-intensive preparation: most of the biographies were adapted from Guardian obituaries. Again there was an attempt to include the unusual to stimulate interest (eg stonemason, Zoroastrian priest, walking to the North Pole).

5. Afterthoughts

No stimulus beyond the note sheet, but the tasks were similar to those in the course book. This simplicity was meant to provoke straight narrative, and students produced it without difficulty. The re-telling of the best story from one group to another resulted in improved language, showing the benefits of rehearsal.

6. Books for the library

A new version of a previous scenario, with a more structured set of tasks, again written directly for students without the intervention of the teacher. The books were the researcher's, and handling them as realia seemed to be a stimulus to student reaction, and the element of speed reading was emphasised by the need to pass on the books to other students.

7. Four characters in search of a drama

A deliberate mix of sculptures, intended to stimulate invention, four (from one corner of the set of 12) to each student in a group, two groups. The class was sitting Cambridge CAE in a few days' time and the teacher welcomed the scenario as practice for the paired oral.

Note

The longer scenarios are represented by samples of the material used.

Appendix 6.2.1

Scenario: Introduction to Dorset

ref ☐

name ☐

Notes

A group of new students are arriving at the school next Friday, but they do not start their courses until Monday. You have been asked to suggest some interesting activities to introduce them to Dorset.

Task 1

The pictures are photographs of the students. Choose one picture and write in the boxes what you think he or she is like:

picture no. ☐

age ☐

nationality ☐

job ☐

interests ☐ ☐ ☐

personality

1 ☐

2 ☐

3 ☐

name ☐

Task 2

Discuss your description with your partner, then choose one activity from the Events List that both the students will be interested in. Write the name of the activity here:

_____

Task 3

As a class, agree on a timetable for the activities day. You can use the 12-seater bus from 0930 until 1830, but you will have to decide where it will go.

Write here what you have agreed for your two new students:

_____

_____

_____

_____

_____

Appendices

Appendix 6.2.1

LAC Assessment Materials: people 1

1

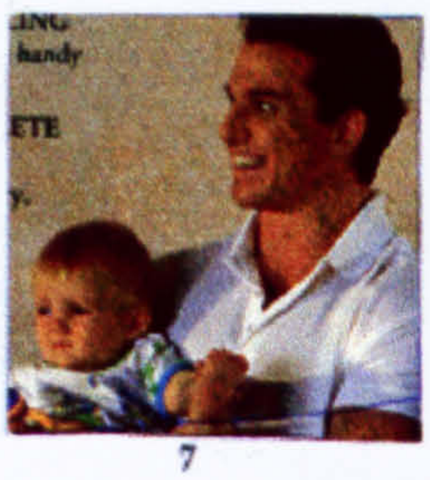
2

3

4

5

6



7

8

9

10

11

12

13

14

15

16

Appendices

Appendix 6.2.1

Appendices

Appendix 6.2.1

Scenario: Introduction to Dorset : Questionnaire [10 mins]

Please tell me what you think about the scenario: put a cross in one box for each question.

1 Class assessment

I think that, as a group, we did	well ☐	fairly well ☐	rather badly ☐

2 Self-assessment

In the scenario,			
I could say what I wanted to say	often ☐	sometimes ☐	not often ☐
I understood and talked	well ☐	quite well ☐	with difficulty ☐
I found that talking about			
- people and relationships was	easy ☐	a problem ☐	difficult ☐
- activities was	easy ☐	a problem ☐	difficult ☐
- other things was	easy ☐	a problem ☐	difficult ☐
- things that are happening now was	easy ☐	a problem ☐	difficult ☐
- things that happened in the past was	easy ☐	a problem ☐	difficult ☐
I found that discussing things with the other students was	easy ☐	a problem ☐	difficult ☐

3 The scenario

This is how successful I think the scenario was:

- it was an enjoyable exercise	yes ☐	not really ☐	no ☐
- it included a lot of what I learned during the course	yes ☐	not really ☐	no ☐
- it showed me what I can do with the language I have learnt	yes ☐	not really ☐	no ☐

4 Notes

Thank you very much for your help.

Andrew Harrison

Appendix 6.2.2

Scenario: A job in Dorset

ref ☐

name ☐

Notes

A friend in Dorset has invited you to stay on after the end of the course. Before you agree, you will need to find work locally so that you will have enough money. What kind of job would make you happy? It could be something different from what you do now.

Task 1

Dorset is mainly farming country and seaside, with lots of people coming on holiday. Think about any interests, hobbies or skills you have which could be useful for a job in the area. Write them here:

Task 2

Discuss your ideas with your partner, then choose one job from the list which you think you could apply for. Write the name of the job here:

Talk with your partner about both your jobs, especially what you would be happy about and what would worry you. When you have talked enough, write notes on the job you have chosen.

happy about: *worry about:*

Task 3

As a class, talk about the most interesting jobs and who is going to apply for them. Try to agree about who is most likely to get the job he or she wants. Make a few notes below on the discussion:

Appendices

Appendix 6.2.2

List of jobs

EAST DORSET GOLF CLUB
HYDE NR WAREHAM
A vacancy exists for a
COOK/CHEF
to join a small brigade producing a high standard of mainly home cooked food.
★ 70/71 (minimum pretered) ★
★ straight shifts but some weekend working ★
★ own transport necessary ★
★ excellent pay and conditions ★
★ Basic Hygiene Certificate an advantage ★
—o—
Telephone Gary Adams 01929-472244 to arrange interview

Part-time TAXI DRIVERS Wanted
Aged 30+, must have clean driving licence.
Hackney Carriage Licence will be required
Telephone (01929) 552495

SWANAGE MIDDLE SCHOOL
Herston, Swanage, Dorset BH19 2PH
LUNCHTIME SUPERVISORY ASSISTANT
Required to work with a team Monday to Friday, 11.45am - 1.05pm daily.
Rate of pay £3.84 per hour.
Please apply in writing to the Headteacher giving names and addresses of two referees.
Closing Date - 30 June 1995

REGISTERED/INSURED
CHILD-MINDER
N.C.M.A.
Now has full part-time vacancies from September
PHONE KERRY 425222

Part-time WASHER-UP
12.30-2.30 pm
★ FIVE DAYS A WEEK ★
£3.18 per hour. Transport provided from Wareham
Tel: (01305) 251888, Ext 2777

WEEKEND PERSON
REQUIRED
Additional hours in week by arrangement. Secretarial experience an advantage
Apply
PALMER SNELL
Estate Agents
The Triangle, Castle Cary, Somerset
Tel: (01963) 351373

RESPONSIBLE cleaners required 7pm - 9pm Monday - Friday, Saturday rota, inter-mount 9am - 10am Saturday 8th July. Mortimers Bakery, Lyte Trading Estate, Yeovil.

TRACTOR DRIVER
required for Frome area, must be experienced, to work own initiative, long hours. Good rates of pay
Tel (01460) 76256

2 ANIMAL ASSISTANTS
required
One to live-in and one to live out, must be fond of cats, near Corfe Castle.
Please call 01929 480474

MUSHROOM Harvester required in Sherborne area, weekday and or weekend work available, suit someone with good hand and eye coordination, own transport essential. Contact George Painting on [01963] 210847 between 9am - 2.30pm.

YEOVIL robotic, automatic Carrier (2CV's) required for Rally season to assist 24 year old man. £4,782 in own name. - Telephone [01963] 340276.

BAKERY assistant required, become interviews 9am - 10am Saturday 8th July. Mortimers Bakery, Lyte Trading Estate, Yeovil.

Appendices

Appendix 6.2.2

Scenario: A job in Dorset : Questionnaire [10 mins]

Please tell me what you think about the scenario: put a cross in one box for each question.

1 Class assessment

I think that, as a group, we did	well ☐	fairly well ☐	rather badly ☐

2 Self-assessment

In the scenario,			
I could say what I wanted to say	often ☐	sometimes ☐	not often ☐
I understood and talked	well ☐	quite well ☐	with difficulty ☐
I found that talking about			
- people and relationships was	easy ☐	a problem ☐	difficult ☐
- jobs was	easy ☐	a problem ☐	difficult ☐
- other things was	easy ☐	a problem ☐	difficult ☐
- things that are happening now was	easy ☐	a problem ☐	difficult ☐
- things that happened in the past was	easy ☐	a problem ☐	difficult ☐
I found that discussing things with the other students was	easy ☐	a problem ☐	difficult ☐

3 The scenario

This is how successful I think the scenario was:

- it was an enjoyable exercise	yes ☐	not really ☐	no ☐
- it included a lot of what I learned during the course	yes ☐	not really ☐	no ☐
- it showed me what I can do with the language I have learnt	yes ☐	not really ☐	no ☐

4 Notes

Thank you very much for your help.

Andrew Harrison

Appendix 6.2.3

Scenario: Compatible friends

ref
name

Notes
Imagine that when you were on holiday you met someone and made friends with them. Now you have received a postcard from this friend, who suggests you should meet again.

Task 1
Read the postcard and think about the person who sent it.

Who sent the postcard? _____

What kind of person is he/she? _____

_____

Task 2

Discuss with your partner what <u>you</u> want to do, and also what <u>your partner</u> wants to do.

Do you want to meet your holiday friend again? _____

If yes, why? if not, why not?

_____

_____

Task 3
The friend will telephone you soon. Discuss with the others in your group what you are going to say. If the answer is 'yes', make arrangements to meet. If the answer is 'no', think of excuses you can give. Write some excuses here to remind you:

_____

_____

_____

Task 4 - Group A

Decide which student in the group will take the phone call. When the phone rings, he or she must answer it and agree (or not agree) to meet the holiday friend, as you have decided!

Task 4 - Groups B & C

One student will make the phone call. Read the notes and use them to help you.

All groups

Write a few notes on what happened.

_____

_____

_____

Appendices

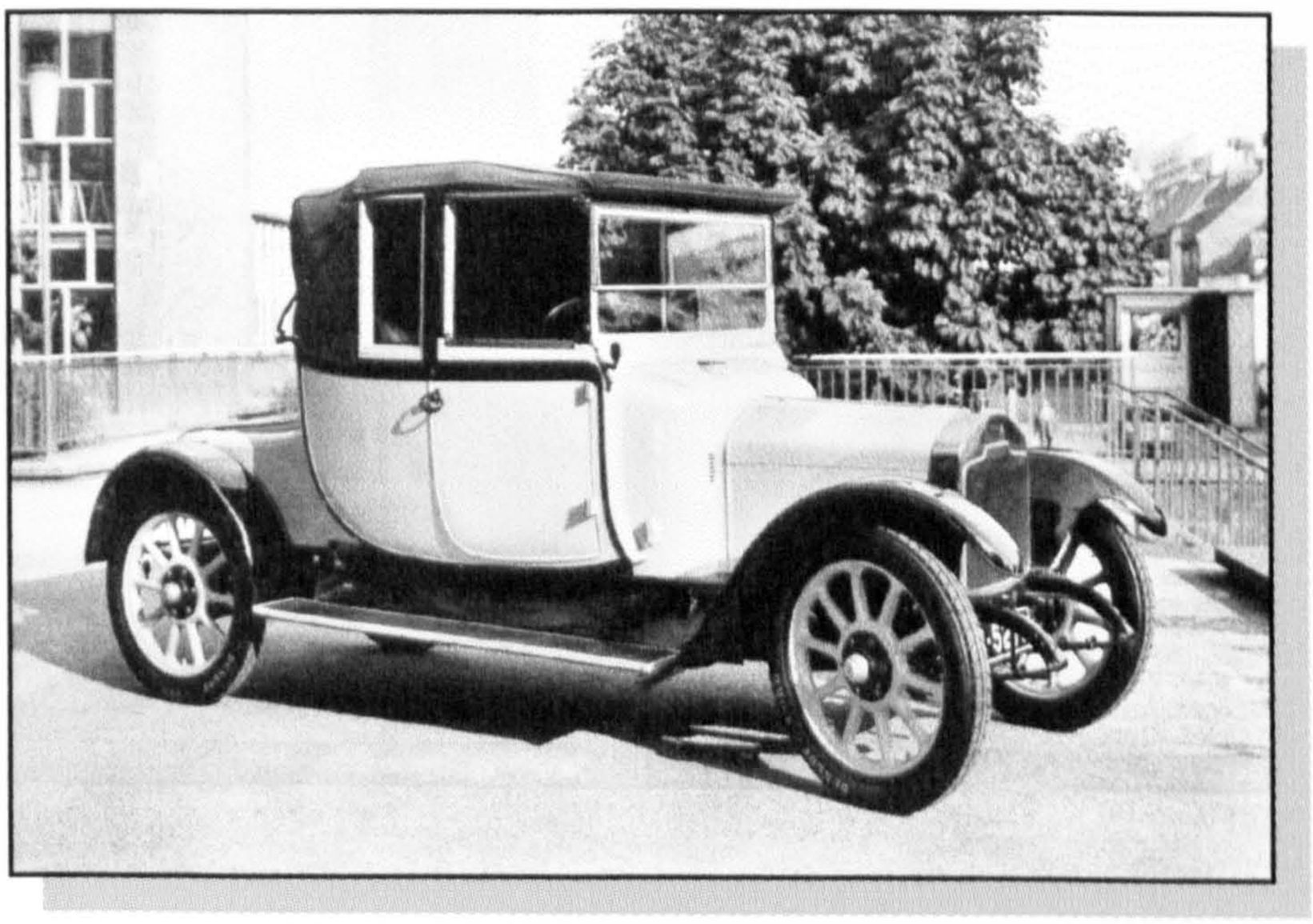
Appendix 6.2.3

It seems a long time since we met on holiday last year. But I have come to England at last and now I'm on my way home again. I'm taking the Channel Ferry from Poole - hope to spend a couple of days walking the Coast Path. I've done the bit along Studland Beach as far as Swanage, but there's much more of it to do - I won't have a lot of time. What about lunch on Sunday at the pub at Worth Matravers? Or would you like to walk with me? I'll ring you on Wednesday. I do hope we can meet again.

Sincerely, Viv

Appendices

Appendix 6.2.3



We said we would meet again one day - & here I am in England again. I hope we can manage it. I've seen some fantastic old vehicles since I came - I'm still going round transport museums, as you see. But I haven't done the Dorset ones yet, especially Beaulieu (though that's in Hampshire). I hope you'll come with me to look at all the wonderful old cars they've got there. You will come, won't you? I'll give you a ring on Wednesday, just as soon as I get to Bournemouth & then we can arrange a whole day going round the museum. I'm looking forward to it! Yours. Jan

Appendix 6.2.3

Viv

Notes for your phone call. Life is difficult, always different from what you expect! Telephone your friend and see what he or she replies.

If they *don't* want to meet you, try the following:
- But you said you were always free [at weekends/on Fridays...].
- But I thought you were interested in walking.
- But I'm only here for two days and I shan't be back in Bournemouth again for years
- But you never told me you lived so far out in the country!

If they *do* want to meet you, try:
- I can't see you on [Saturday/ Sunday...], I have to go to a meeting - I'm terribly disappointed.
- I'm sorry - I'm ringing to tell you that I really can't see you after all - my [husband/wife/partner...] has just come over to join me.
- The trouble is, I've hurt my ankle, so I can't walk very far.
- I'm ever so sorry, but I can't stay in Bournemouth after all - my mother's desperately ill.

Jan

Notes for your phone call. Life is difficult, always different from what you expect! Telephone your friend and see what he or she replies.

If they *don't* want to meet you, try the following:
- But I thought you liked old cars as much as I do.
- But you said you were always free [at weekends/on Fridays...].
- But it's not very far over to Beaulieu - only about half an hour.
- But I do need a whole day, otherwise I won't see everything.

If they *do* want to meet you, try:
- I can't see you on [Saturday/ Sunday...], I have to go to a meeting.
- I'm very disappointed, but the car museum is closed for repairs.
- I'm sorry, but I find my flight is booked for tomorrow and I can't change it.
- I'm ever so sorry, but I can't stay in Bournemouth after all - my mother's desperately ill.

Appendix 6.2.3

Scenario: Compatible Friends : Questionnaire [10 mins]

Please tell me what you think about the scenario: put a cross in one box for each question.

1 Class assessment

I think that, as a group, we did well ☐ fairly well ☐ rather badly ☐

2 Self-assessment

In the scenario,

I could say what I wanted to say	often ☐	sometimes ☐	not often ☐
I understood and talked	well ☐	quite well ☐	with difficulty ☐
I found that talking about			
- people and relationships was	easy ☐	a problem ☐	difficult ☐
- arrangements & excuses was	easy ☐	a problem ☐	difficult ☐
- other things was	easy ☐	a problem ☐	difficult ☐
- things that are happening now was	easy ☐	a problem ☐	difficult ☐
- things that happened in the past was	easy ☐	a problem ☐	difficult ☐
I found that discussing things with the other students was	easy ☐	a problem ☐	difficult ☐

3 The scenario

This is how successful I think the scenario was:

- it was an enjoyable exercise	yes ☐	not really ☐	no ☐
- it included a lot of what I learned during the course	yes ☐	not really ☐	no ☐
- it showed me what I can do with the language I have learnt	yes ☐	not really ☐	no ☐

4 In comparison

Compared with the one we did before, this scenario was better ☐ about the same ☐ worse ☐

5 Any further comments

_____

_____

_____

_____

Thank you very much for your help.

Andrew Harrison

Appendix 6.2.4

Scenario: Lecturer ref
name_____

Notes
The school arranges a lecture on Monday and Thursday evenings. In a good democratic way, it has been decided that you, the students, should decide who is to be invited to come and give a lecture for everybody.

Task 1
What do you think are the most important qualities for a lecturer? Here are three:
an outgoing personality ... something interesting to say ... a sense of humour ...
Put these in order of importance by writing 1, 2 or 3 beside them, then add three more of your own:

_____

_____

_____

Task 2
Now look at the information about three people on Sheet 1, *and with your partner decide which <u>two</u> people you would like to ask to give the lecture, from the six people on your two sheets.*
1_____ 2_____
Why did you choose these two?

_____

_____

_____

Task 3
Read the information you now have on Sheet 2 *about the people you have chosen and discuss in the group which two are now the best.*
1_____ 2_____

Task 4
Now as a class try to make a final choice of lecturer. If you cannot agree on one name, you may suggest two. Then, on your own, write a few notes on the discussion.

_____

_____

_____

_____

_____

Appendices

Appendix 6.2.4

name	*age*	*occupation*	*can talk about*	*aim in life*	*strengths*
Anna Adamson	32	editor	preparing authors' writing for publication	helping people	understanding
Brian Booth	35	sports coach	his training system	improve athletes' performance	confidence
Charles Carter	78	doctor (retired)	his hospital in South Africa	healing	dedication
David Dench	28	stonemason	carving in stone, mainly for buildings	accuracy	craftsmanship
Ed Erlanger	61	historian, author	post-industrial culture about society	find the truth think in US	ability to think
Francis Ford	35	explorer	walking to the North Pole	find out about the world	determination
Georgina Garton	82	teacher	secondary education in England	the best education for each pupil	efficiency
Hari Harishna	56	Zoroastrian priest	history of religion	continuing old traditions	always tells the truth
Ian Inglis	70	adventurer	escaping from prison camps	always being right	single-mindedness
John Jackson	57	writer	discussing social problems (eg prejudice)	writing & being published	sense of purpose
Keith Kendal	17	trainee warden, National Trust	care of the countryside forestry	become a warden in Scotland	enthusiasm
Lesley Laughton	?	dancer, choreographer	the human body & movement	expressing emotion in dance	glad to be her own master

Appendix 6.2.4

Anna Adamson
She is a wonderful editor - writers discreetly hand Anna their copy and, with a bit of judicious cutting here and there, a decent piece is transformed into a very good one. Everyone goes to her for comfort. Writers phone up, bitterly asking why their article has not been used. Anna will talk them through it, highlight the best bits, make them up if necessary, and by the time she has finished they feel like God's gift to journalism. But she is always convinced she looks a mess.

(adapted from *The Guardian*, 17/5/95)

Brian Booth
'I am a scientist,' he says, 'and when I was running myself, and having bad days and good days, I thought, there's got to be a reason for this, and it's got to be predictable. I spent five years researching it.' Broadly, his method is similar to bio-rhythms, the theory that immutable, innate body rhythms - such as the daily 24-hour cycle and the lunar monthly and yearly cycles - combine with changeable cycles, dependent on human habits and character, to produce a three-mode system: the intellectual cycle (33 days), the emotional cycle (28 days) and the physical cycle (23 days).

(adapted from *The Guardian*, 14/8/93)

Charles Carter
Charles and Clara Carter arrived in South Africa in 1945 as medical missionaries to take over a dilapidated store in Zululand. Twenty-seven years later, when forced by their principles to leave - they could not countenance a government takeover - they had built up not only a large, well-equipped hospital, which drew doctors from all over the world, but also a deeply sympathetic relationship with the community, winning the confidence and respect even of traditional healers. Since their retirement they have cycled tens of thousands of miles on their tandem, raising money for charities in England and South Africa.

(adapted from *The Guardian*, 4/9/93)

David Dench
David's work is extremely varied, ranging from repairs to local church pinnacles and carving capitals for new buildings for the City of London, to making garden benches. Much of his masonry is created at home. In an idyllic setting behind the house is a small stone barn, which he has turned into his workshop. Everything inside is thick with stone dust. Machinery is noticeable by its absence. 'I use the same type of tools that stonemasons have used for hundreds of years,' he says. David's work is meticulous, often incredibly detailed and almost painfully slow.

(adapted from *Homes & savings*, Halifax Building Society, Spring 1991)

Ed Erlanger
Erlanger's books analyse modern industrial consumer society. He argues that we think we now have limitless options without consequences, the 'freedom to choose everything at once'. We define who we are through purchases, through opinion polls - reactions to pre-set agenda. We have almost no control over our world. The communications industry channels our information, politics studies and manages our opinions. No one has purpose, or faith in the future: all that can be done is to chill off, be ironic. But he has some hope for a post-industrial culture that will accept limits, respect nature, be humble, local, traditional... and self-disciplined.

(adapted from *The Guardian*, 19/2/94)

Appendix 6.2.4

Scenario: Lecturer : Questionnaire [5 mins]

Please tell me what you think about the scenario: put a cross in one box for each question.

1 Class assessment

I think that, as a group, we did	well ☐	fairly well ☐	rather badly ☐

2 Self-assessment

In the scenario,			
I could say what I wanted to say	often ☐	sometimes ☐	not often ☐
I understood and talked	well ☐	quite well ☐	with difficulty ☐
I found that talking about			
- people personalities was	easy ☐	a problem ☐	difficult ☐
- arrangements & excuses was	easy ☐	a problem ☐	difficult ☐
- other things was	easy ☐	a problem ☐	difficult ☐
- things that are happening now was	easy ☐	a problem ☐	difficult ☐
- things that happened in the past was	easy ☐	a problem ☐	difficult ☐
I found that discussing things with the other students was	easy ☐	a problem ☐	difficult ☐

3 The scenario

This is how successful I think the scenario was:

- it was an enjoyable exercise	yes ☐	not really ☐	no ☐
- it included a lot of what I learned during the course	yes ☐	not really ☐	no ☐
- it showed me what I can do with the language I have learnt	yes ☐	not really ☐	no ☐

4 In comparison

Compared with the one we did before,
this scenario was better ☐ about the same ☐ worse ☐
Why?

_____

_____

5 Any further comments

_____

_____

Thank you very much for your help.

Andrew Harrison

Appendix 6.2.5

Scenario: Afterthoughts ref:

name_____

Notes

The editor of the BEET newsletter is asking for contributions from students. This class has been asked to write a short story. This is one way you could do it.

It is difficult sometimes to do the right thing at the right time. How often have you regretted saying something, or not saying something, and then thought what you should have said? Or perhaps it was something you did and then wished you hadn't?

Task 1

Think of an event or conversation in which you wish you had done or said something different. Write notes about it here:

What happened was _____

If it happened again I would _____

Task 2

Working in a group of four, tell your story to the others and listen to theirs. Write a few words to remind you what they said:

1 _____

2 _____

3 _____

Best story _____

Task 3

The student who told the best story takes it to another group and this group decide whether the new story is better than their own story, or not so good.

The better story is ours... theirs...

Task 4

As a class, decide which story is to be sent in to the editor.

Appendix 6.2.5

Scenario: Afterthoughts : Questionnaire [5 mins]

Please tell me what you think about the scenario: put a cross in one box for each question.

1 Class assessment

I think that, as a group, we did well ☐ fairly well ☐ rather badly ☐

2 Self-assessment

In the scenario,

I could say what I wanted to say	often ☐	sometimes ☐	not often ☐
I understood and talked	well ☐	quite well ☐	with difficulty ☐
Speaking quickly enough was	easy ☐	a problem ☐	difficult ☐
I found that talking about			
- what actually happened was	easy ☐	a problem ☐	difficult ☐
- what I wished had happened was	easy ☐	a problem ☐	difficult ☐
- things that are happening now was	easy ☐	a problem ☐	difficult ☐
- things that had happened in the past was	easy ☐	a problem ☐	difficult ☐
- things that may happen in future was	easy ☐	a problem ☐	difficult ☐
I found that discussing things with the other students was	easy ☐	a problem ☐	difficult ☐

3 The scenario

This is how successful I think the scenario was:

- it was an enjoyable exercise	yes ☐	not really ☐	no ☐
- it included a lot of what I learned during the course	yes ☐	not really ☐	no ☐
- it showed me what I can do with the language I have learnt	yes ☐	not really ☐	no ☐

4 Notes

Thank you very much for your help.

Andrew Harrison

Appendix 6.2.6

Scenario: Books for the library ref:

name _____

Notes

Someone has given a lot of old books for the school library, but there are far too many of them. As students in an advanced class, you have been asked to choose the best ones. Here are 8 books for you to look at:

	your comments
Adams: Watership Down	_____
Bach: Jonathan Livingston Seagull	_____
Chatwin: Utz	_____
Godden: The River	_____
Greene: Brighton Rock	_____
Hemingway: The Old Man and the Sea	_____
MacLean: Where Eagles Dare	_____
Spielberg: Close Encounters...	_____

Task 1

Work with a partner. When you are given a book, read the blurb (the information about it which is printed on the back, or sometimes inside at the front), and then look quickly at the beginning and the end of the book itself. Discuss with your partner what you think about it, then write a few comments beside the title in the list above. You should take no more than 3 minutes to look at each book. Now do the same with the other seven books.

Task 2

Working on your own, decide which of the books you think would be the best three additions to the library.

1 Best book: _____

Suitable for: younger students ☐ adults ☐ elementary ☐ intermediate ☐ advanced ☐

What made you choose it? _____

2 Second book: _____

Suitable for: younger students ☐ adults ☐ elementary ☐ intermediate ☐ advanced ☐

What made you choose it? _____

3 Third book: _____

Suitable for: younger students ☐ adults ☐ elementary ☐ intermediate ☐ advanced ☐

What made you choose it? _____

Task 3

As a class, discuss what you have recommended and put all the books in order of preference.

Appendix 6.2.6

Scenario: Books for the library : Questionnaire [5 mins]

Please tell me what you think about the scenario: put a cross in one box for each question.

1 Class assessment

I think that, as a group, we did well ☐ fairly well ☐ rather badly ☐

2 Self-assessment

In the scenario,			
I could say what I wanted to say	often ☐	sometimes ☐	not often ☐
I understood and talked	well ☐	quite well ☐	with difficulty ☐
Reading quickly enough was	easy ☐	a problem ☐	difficult ☐
I found that talking about			
- the contect of the books was	easy ☐	a problem ☐	difficult ☐
- what nakes a good book was	easy ☐	a problem ☐	difficult ☐
- things that are happening now was	easy ☐	a problem ☐	difficult ☐
- things that happened in the past was	easy ☐	a problem ☐	difficult ☐
- things that may happen in future was	easy ☐	a problem ☐	difficult ☐
I found that discussing things with the other students was	easy ☐	a problem ☐	difficult ☐

3 The scenario

This is how successful I think the scenario was:

- it was an enjoyable exercise	yes ☐	not really ☐	no ☐
- it included a lot of what I learned during the course	yes ☐	not really ☐	no ☐
- it showed me what I can do with the language I have learnt	yes ☐	not really ☐	no ☐

4 Notes

Thank you very much for your help.

Andrew Harrison

Appendix 6.2.7

Scenario: Four characters in search of a drama

ref:
name_____

Notes
The end of term show at the school, UPBEET, includes performances from students - songs, dances or anything which contributes to the entertainment. It has been suggested that this class should present a short play. It can be any kind of drama - a serious story, a comedy, a dream, a myth, some kind of science fiction... We will start with the characters.

Task 1

Which one of the people in your pictures would you like to be, as a character in a play? No.__

What is his or her name? (Invent one if you need to!) _____

What sort of person is he or she? Find four suitable words and draw a ring round each of them.

brave tender inscrutable careful idealistic dreamy ambitious graceful pensive energetic aggressive motherly superior strong secretive hardheaded anxious powerful caring arrogant forceful neat mysterious weak unfriendly sensitive warm relaxed cold

Add three more descriptions here: _____

1_____ 2_____ 3_____

What is the one most important thing about this character?

_____

Task 2

Work with a partner. Take turns.
a. One of you gives the other his or her four pictures, then describes the character, asking which picture it is. ***Do not say what the character looks like****, just explain what sort of person the character is.*
b. Then the other partner looks at the other four pictures and decides from the character description which one it is.

Task 3

Working in a group of four, invent a short play in which the four characters are involved. Where do they meet? What do they say to each other? What happens?
If it helps, write few notes on the play here:

_____

_____

_____

Task 4

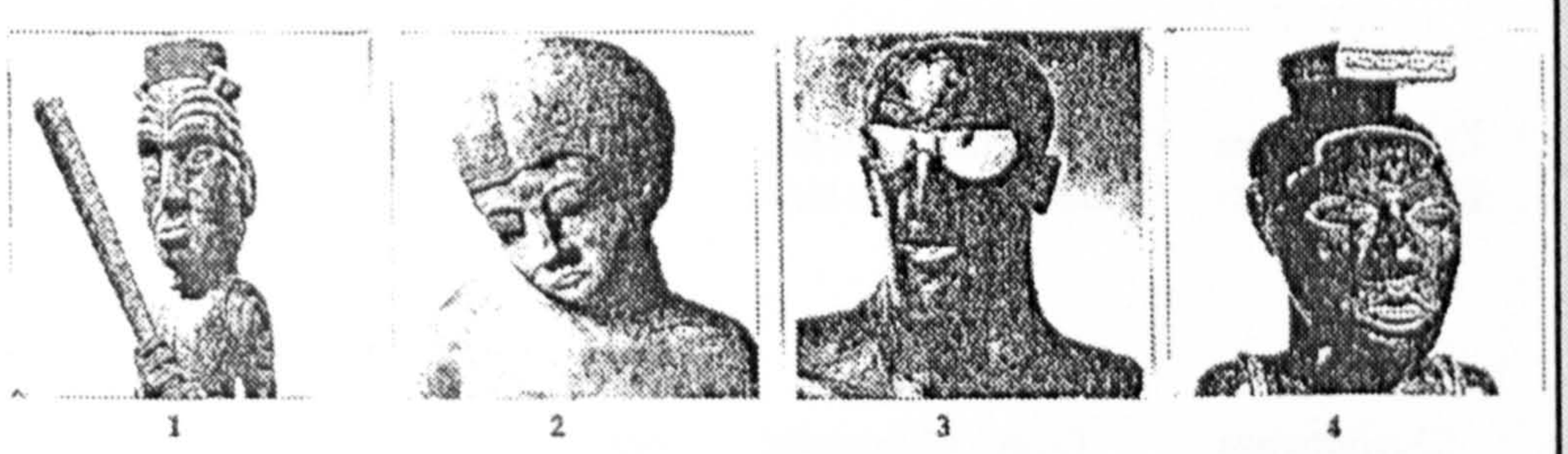
Tell the other group what your play is about. Act it for them, if you like.

Task 5

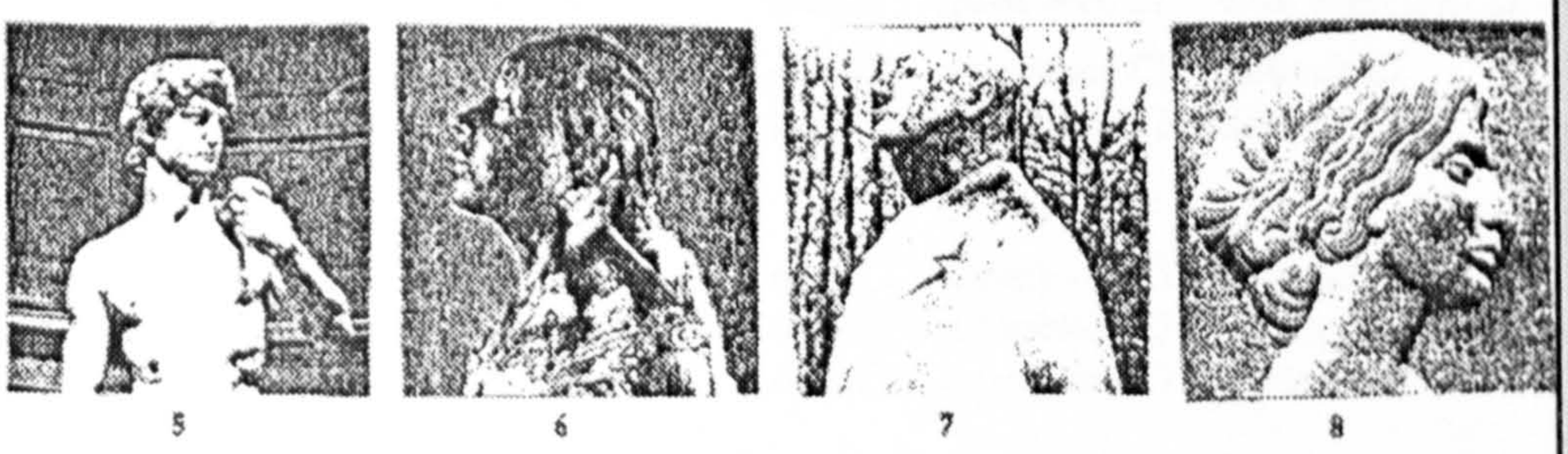
As a class, decide which of the two plays is the one you will perform at the end of term concert.

Appendix 6.2.7

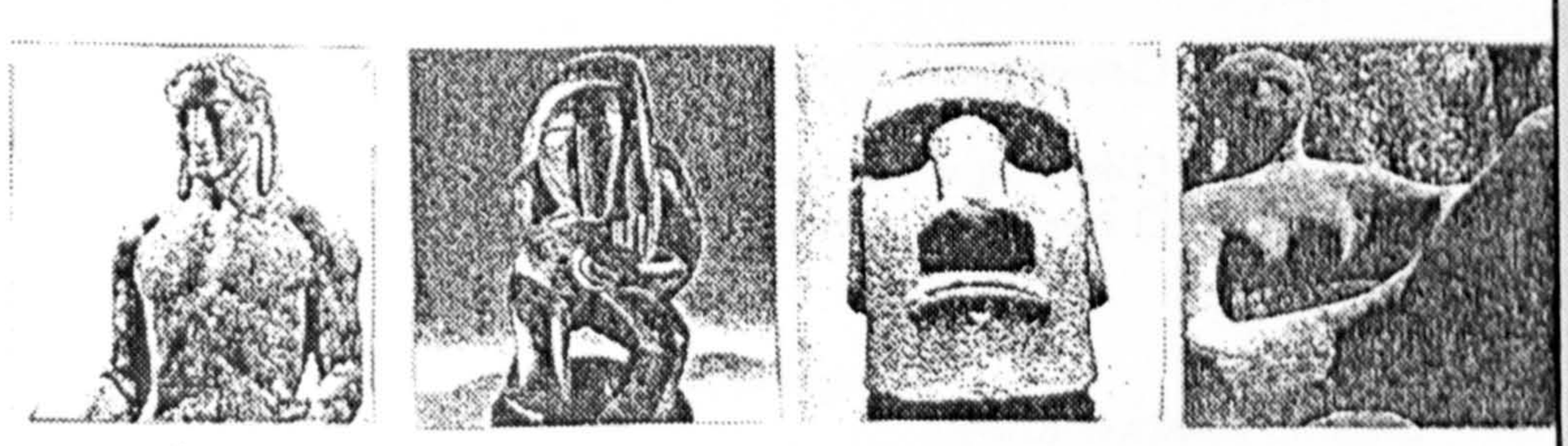
LAC Assessment Materials: people 2



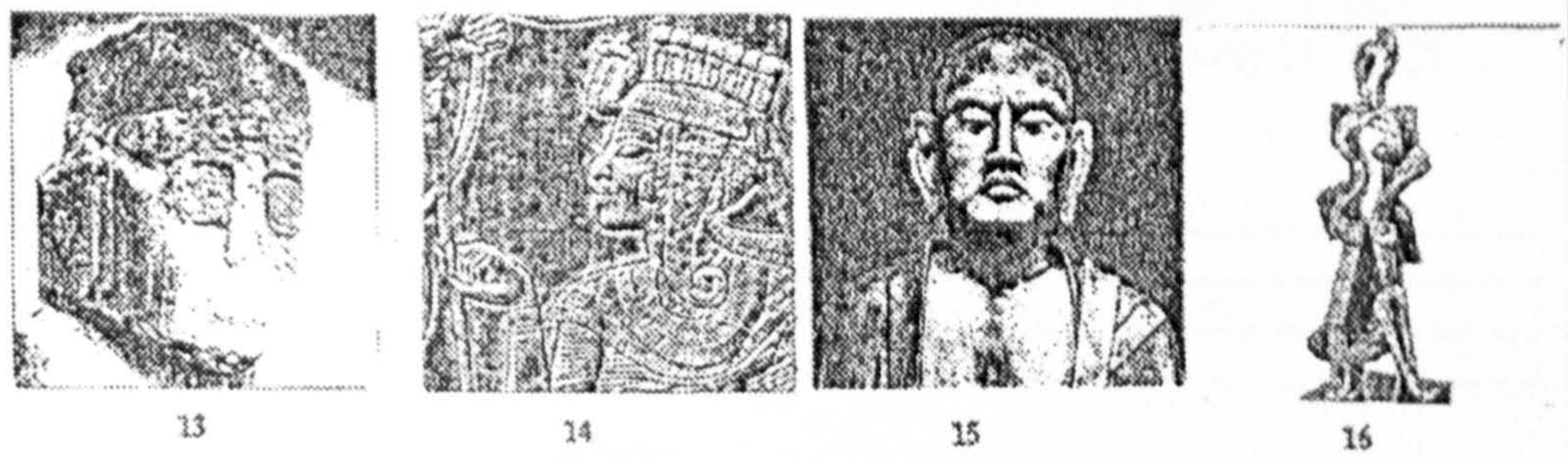
1 2 3 4



5 6 7 8



9 10 11 12



13 14 15 16

Appendix 6.2.7

Scenario: Four characters : Questionnaire [5 mins]

Please tell me what you think about the scenario: put a cross in one box for each question.

1 Class assessment

I think that, in our group, we did	well ☐	fairly well ☐	rather badly ☐
I think that, as a class, we did	well ☐	fairly well ☐	rather badly ☐

2 Self-assessment

In the scenario,			
I understood what people said to me	always ☐	often ☐	sometimes ☐
I could say what I wanted to say	well ☐	quite well ☐	with difficulty ☐
To describe a character from a picture was	easy ☐	a problem ☐	difficult ☐
To recognise my partner's character was	easy ☐	a problem ☐	difficult ☐
I found that talking about			
- the story of the play was	easy ☐	a problem ☐	difficult ☐
- the characters in the play was	easy ☐	a problem ☐	difficult ☐
- the end of the play was	easy ☐	a problem ☐	difficult ☐
I found that discussion with the other students was	easy ☐	a problem ☐	difficult ☐
The grammar I needed was	easy ☐	a problem ☐	difficult ☐
The vocabulary I needed was	easy ☐	a problem ☐	difficult ☐

3 The scenario

This is how successful I think the scenario was:

- it was an enjoyable exercise	yes ☐	not really ☐	no ☐
- it included things I learned during the course	yes ☐	not really ☐	no ☐
- it was useful practice for the exam	yes ☐	not really ☐	no ☐

4 Notes

_____

_____

_____

_____

Thank you very much for your help.

Andrew Harrison

Appendices

Appendix 7.1.1

A mark scheme in action: two learner texts

The scenario used for this exercise was 'Books for the library', which asks students to work together in pairs (and later in groups) to decide which of eight old books are the best three to be taken into the school library. The texts have been adapted from transcripts of an hour's discussion in class. The conventions followed in the transcripts are: no punctuation supplied; space indicates pause; square brackets indicate comments; ... implies a break in continuity.

Text 1

This text represents a discussion between B1 and B2 about the books. They have been told that they have about three minutes to look at each of a series of books and decide whether they think it is for adults or for younger students, and what level of language the book demands - elementary, intermediate or advanced. It has been suggested that to help them, they should look at the information given for each book on the cover and inside it.

		interaction	*resources*				*activities*	
			s	*/*	*s*	*/*		
			B1	B1	B2	B2	B1	B2
01	B2 I hope it was	-						
02	we've got three more to go	init			✓	✓		k
03	haven't got much time	cont/offer			✓	✓		k
04	[blurb] 'truth to character and situation is the real attraction of the novel should be read by everyone beautiful writing and delicate'	init			✓			u
05	think actually it should be for adults who can appreciate much more than just you know one step at a time	cont/offer			✓	✓		p
	...							
06	B2 what else can we get out of this	init			✓	✓		k
07	B1 wait a minute this is greatest information about this	-						
08	B2 yeah but	resp				✓		
09	B1 [blurb] to catch the spirit of this perfectionist novel	cont	✓					
10	B2 but this is for Mermaid Books about the series it is not about this special book this group of books what they represent	resp/denial			✓	✓		p
11	B1 [continues reading blurb sotto voce]	-						
12	B2 all right so they are saying that these may be cheap books but people appreciate them as they are interesting	resp			✓	✓		u
	...							

		interaction	*resources*				*activities*	
			s B1	*l* B1	*s* B2	*l* B2	B1	B2
13	B2 I think it's for youngers about all these youngers and confidences ...	init				x		p
14	B1 it's the young spirit of the books which shows that	-					u	
15	B2 I mean it's supposed to be like the theme of the book is life	cont			✓	✓		u
16	so I should say it's for adults	cont			✓	✓		p
17	I mean you get adults from let's say 25 up to 80 years old	cont			✓	✓		u
18	It's very difficult to decide but I should say that this is not for children	cont/offer			✓	✓		p
19	but it is like for for young persons	cont			✓	✓		p
20	B1 I mean young persons this will be why they set the price	cont/offer	✓	✓			u	
21	B2 yes	resp				✓		
22	B1 and that they had students in mind	cont/implic	✓	✓			u	
23	and I'd say the problem for us is the real appreciation of the level	offer	✓	✓✓			u	
	[pause while they read]							
24	B2 so it's actually quite intelligent book	init			x			u
25	B1 good writing	cont	✓				u	
26	B2 I mean for intelligent people so so it has to be what stage?	cont/enq			✓	✓		r
27	advanced I suppose	cont/offer			✓	✓		p
28	so I would say for adults and advanced language	concl			✓	✓		p
29	B1 but a young person's story	resp/denial		✓				
30	B2 I think it is for adults I would say this is for adults	cont			✓	✓		p
31	this is like for everyone who'd like to be remind of childhood	cont/implic			x	✓		u
32	B1 mm yes I'd say so	resp	✓	✓			r	
	...							
	[pause while they look at next book - *Brighton Rock*]							
33	B2 OK I would say that this is advanced because I consider Penguin talking about er the production whatever it is who publisher publisher	init/aid			✓	✓		u
34	so they are they are having their collection in order to	-						
35	B1 quite classic collection	init		✓			u	
36	B2 yes or this one or that one *Buddenbrooks* [ad in book for another in the 'classics' series] this is quite difficult	cont/offer			✓	✓		p
37	I would say that this is for advanced advanced	cont/offer			✓	✓		p
38	B1 advanced students	resp	✓				r	
39	B2 from the story definitely not for children	cont				✓	u	
40	B1 mm	resp		✓				
41	B2 so its about a boy of 17	init			✓	✓		u
	[reads sotto voce]							
42	B1 yeah this is definitely higher level	cont		✓			p	
	...							

		interaction	resources s B1	resources l B1	resources s B2	resources l B2	activities B1	activities B2
43	B2 [blurb] 'a mystery story adventure psychological action-story at least it ... a novel which when it is finished seems to have been injected into the veins'	cont				✓		u
44	B1 yes it seems action structure novel	cont	x	✓			r	
45	B2 quite difficult	init				✓		p
46	B1 it is difficult?	cont/enq	✓	✓			r	
47	B2 it is I think it is quite difficult yeah	resp			✓	✓		p
48	I would say that this I mean this is psychological action story so you wouldn't give it to children would you? a psychological book	cont/offer			✓✓	✓		p
49	B1 no certainly no	resp	✓				r	
50	B2 they wouldn't be able to understand it	cont			✓	✓		u
51	so I think this is definite quite definitely advanced language for adult people	concl			✓	✓		p
52	and you know for someone who like classics basically like *Buddenbrooks* or you know	cont/implic			x	✓		
53	B1 yes	resp	✓				r	
54	B2 I mean just just read this bit in here first of all	cont			✓	✓		u
55	[blurb] 'deep mystery and yet it is not mystery story there is rare adventure yet more than adventure tale psychological action-story mood scene and character a novel which when it is finished seems to have been injected into the veins'	cont				✓		u
56	B1 yeah but concerning youngsters they have more excitement and adventure and and so on with films	resp/denl	✓	✓			u	
57	B2 but I think you have to if we are talking about younger students at the age of 15 12-15 they wouldn't be able to understand this book	resp/denl			✓	✓		u
58	B1 |yeah not if they read outside the teaching	resp	✓	✓				
59	B2 |because the action of the story I mean it's different like you know different things from 20 at about 20	cont/offer			✓	✓		u
60	B1 yeah adults	resp		✓			r	
61	B2 so this sort of book is basically for adults ...	concl			✓	✓		p
62	B2 so shall we go for another one? we have two more	init			✓	✓		k
63	B1 *Where Eagles Dare*	init		✓				
64	B2 it looks like sort of second world war	cont/offer			✓	✓		u
65	B1 germanic	cont/offer		✓				
66	B2 germanic	resp				✓		r
67	B1 this man could write a very good story	init	✓	✓			u	
68	B2 a British expedition or is it ?	init				✓		
60	um I would say this is something for adventure readers only	cont/offer			✓	✓		
70	and I would say that this is for adults	cont/offer			✓	✓		
71	and I would say this is intermediate so not very high level ... well one or two hard bits hard bits you know	cont/offer			✓	✓✓		p
72	an expedition something's gone wrong an action like action book	cont/offer			✓	✓		u

		interaction	resources s B1	resources l B1	resources s B2	resources l B2	activities B1	activities B2
73	not very important	cont/offer				✓		u
74	I would say something that you are reading just you know make your blood pressure become higher	cont/offer				x		u
75	B1 [laughs]	resp		✓				
76	actually we don't know which level they must understand	init	✓	✓			p	
77	reading pieces will give us idea of which language level	cont/offer	✓	✓			r	
78	B2 OK I would say that this is sort of action story because all this from this serial er this cover what I was saying about it	init			x	✓		u
79	B1 but it's something that's exciting	init	✓	✓			u	
80	B2 yes yes its easier than classical	resp			✓	✓		r
81	about murders I don't know about	cont			✓			
82	B1 action	resp/aid		✓			u	
83	B2 action a secret conference	aid/offer				✓		r
84	so I would say this is intermediate of	init				✓		p
85	er I mean good for men not for women	init				✓		p
86	B1 [laughs]	resp		✓				
87	B2 not sort of book for the pillow yes so I could say this was not my culture ...	init/e				✓		u
88	B2 yes sort of like there is romance and there is there is	init				✓		u
89	B1 suspense and hero	cont/offer	✓				r	
90	B2 OK nothing that I could put as a top book	concl				✓		p
91	B1 no	resp	✓				r	

[they write on note sheets]

Appendices

Appendix 7.1.2

Text 2

This text represents a group dicussion among the four students who were previously in pairs and are now talking about their suggestions for which books are to be chosen, whether for adults or younger learners and at what level.

		interaction	resources								activities			
			s	l	s	l	s	l	s	l				
			A1	A1	A2	A2	B1	B1	B2	B2	A1	A2	B1	B2
01	A1 ... let me say about one thing we wanted to put the same work as well as you did	init	✓	x							p			
02	but the first book is one that I saw in the programme you have to read it so not in the library	cont		✓							p			
03	B2 ... if you have to read it its better to make it easy to find it somewhere it makes it more	cont/offer							✓	✓				p
04	so that's why we chose it because it's classic so everyone should read it and everyone should have practice with this book ...	cont								✓				p
05	A2 I do agree ...	resp				✓								
	[next book]													
06	A1 Er we chose MacLean *Where Eagles Dare*	init	✓	✓							p			
07	for it's suitable for younger children	cont	✓	✓							p			
08	and we think that it's a nice adventure... it's a clever action novel	cont	✓	✓							u			
09	it's talking about er war which is a very important matter we think it's very educational very important for children to know about war	cont/e	✓	✓							p			
10	A2 for younger students which helps them to learn the language	cont/offer	✓	✓							p			
11	this is the book to educate you which is a history book with facts which we cannot change and you learn ...	cont	✓	✓							p			
12	B2 but it can be action book for adults for men who will then look for a chance to find excitement	resp							✓	✓				u
...														
13	B2 anyway our second book is *Jonathan Livingston Seagull*	init							✓	✓				p
14	and it's for er we decided actually this is													

		interaction	resources								activities			
			s A1	l A1	s A2	l A2	s B1	l B1	s B2	l B2	A1	A2	B1	B2
	for adults and it it's not very difficult like intermediate level and er	cont							✓	✓				p
15	A2 no no no	resp/den				✓						p		
16	B2 and we decided that we would put it in	cont							✓	✓				p
17	because it's very ... so it's good for students it's quite	cont							✓					
18	the language used for to put on a piece of paper something with a lot of thought behind so even if you people would know the language but you have to think what you're reading	cont/impli							x	✓				p
19	A1 that's what makes it difficult that simple words make such difficult meaning	resp		✓	✓						u			
...	[several voices together]													
20	B2 you can read all the language because it's easy and you have to use your brain to find out what it's about	cont							✓	✓				u
21	B1 ... easy to read but to make you think about it	cont					✓	✓					u	
22	A2 that is true when I read this book I found the thinking difficult	resp			✓	✓						u		
23	I read this book three times I mean if it was really readable I wouldn't have to read it three times	cont			✓✓	✓						u		
24	the vocabulary isn't difficult but the content makes you thinking a lot about	cont			x	✓						u		
25	and really the message and the contents you must read it er many times	cont/offer				x						u		
26	A1 we didn't think it really for the young students to buy specially ...	resp		✓	✓						p			
27	B2 but that's what I mean it's actually easy to read as a as a book because there are simple words	cont/offer							✓	✓				u
28	but there's a lot of thought so that's why it is not for children that's why it is good for adults	cont							✓	✓				p
29	A2 I think that's one of the reasons because if you're mature enough	resp			✓	✓						u		
30	A1 maybe reading is good but the understanding isn't there	resp		✓	✓						u			
	[third choice of book]													
31	A2 *Watership Down* for intermediate students and young readers easy to read novel fiction and educational book	init				✓						u		
32	B2 we put [in third place] Greene - it's *Brighton Rock* and we put something something classical [laughs]	init							✓	✓				p
33	we thought it's for adults for advanced language quite difficult to read	cont/offer							✓	✓				p
34	but it's for entertainment so er instructional book	cont								x				p
35	A2 we just thought that a good book													

		interaction	resources								activities			
			s A1	l A1	s A2	l A2	s B1	l B1	s B2	l B2	A1	A2	B1	B2
	we liked it I mean for its a good book for leisure erm	resp			✓	✓						P		
36	A1 entertainment	aid		✓							u			
37	A2 and entertainment	aid				✓						u		
38	A1 it's not really the book a library needs if you want to you can read it on your own	cont/offer	✓	✓							u			
39	A2 yeah	resp				✓						u		
40	B2 yeah but I don't want to buy that book	resp					✓	✓						p
	[laughter]													
41	A1 that's why I would have so many books for everybody's choice	init	✓	✓								P		
42	B2 yes it's just difficult to choose three books	resp							✓	✓				r
43	A1 no we get	-												
44	B2 it's almost impossible because if if everyone would have different opinion everyone would like...	resp							x	✓				r
	... [several voices together]													
45	B2 one person would prefer to read Shakespeare another one would prefer to read children's stories because it's just easier they want to do light read not to use brain to think about what they are reading	cont/offer								✓				r